

November 2005

Published_Applications Nucleic Acid and Published_Applications Amino Acid database searches now generate two sets of results each. The Published_Applications databases have been split into two parts to reduce the amount of time required for their daily updates. This results in more machine time being available for processing searches.

Newly published applications will appear in the Published_Applications_New databases; older published applications make up the Published_Applications_Main databases.

Searches run against Nucleic Acid Published_Applications produce two sets of results, with the extensions **.rnpbm** (Published_Applications_NA_Main) and **.rnpbn** (Published_Applications_NA_New). Searches run against Amino Acid Published_Applications produce two sets of results, with the extensions **.rapbm** (Published_Applications_AA_Main) and **.rapbn** (Published_Applications_AA_New).

THIS PAGE BLANK (USPTO)

Db 400 EPEAELEERAKSAKPP--QSSSTEPARKP-GQKEKRVPEEKQQA-KPVKVERT-RKRS 454
QY 272 EAAAAKEKNSKQSRVSTD 289
Db 455 EGFSDMRKVEKKEPSVE 472

RESULT 2

US-11-135-855-29
; Sequence 29, Application US/11135855
; Publication No. US2005025557A1
; GENERAL INFORMATION:
; APPLICANT: SMITHKLINE BEECHAM CORPORATION
; TITLE OF INVENTION: NOVEL COMPOUNDS
; FILE REFERENCE: GP50013
; CURRENT APPLICATION NUMBER: US/11/135,855
; PRIOR FILING DATE: 2005-05-24
; PRIOR APPLICATION NUMBER: US/10/203,708
; PRIOR FILING DATE: 2002-08-13
; PRIOR APPLICATION NUMBER: PCT/US01/04703
; PRIOR FILING DATE: 2001-02-14
; PRIOR APPLICATION NUMBER: 60/182,172
; PRIOR FILING DATE: 2000-02-14
; PRIOR APPLICATION NUMBER: 60/186,084
; PRIOR FILING DATE: 2000-02-29
; NUMBER OF SEQ ID NOS: 46
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 29
; LENGTH: 717
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-135-855-29

Query Match 6.3%; Score 143.5; DB 7; Length 717;
Best Local Similarity 24.8%; Pred. No. 0.003;
Matches 79; Conservative 56; Mismatches 138; Indels 45; Gaps 15;

QY 2 AFQSSPLVKFASPAESFLS-APGDNFTSLFADSTP-----STLNPRDMWTPD 49
Db 211 ARKASDDQASVSPSEENSESSSEKTSQDFTPEKKAAVRAPRRGPLGGRKKKKAP 270
QY 50 SVADIDSRL---SVIPESQDAEDDESHSTSATPSTSE---KKPVKKRKSQGVLPPEPKT 103
Db 271 SASDSKADSDGAKPEPVAMARSASSSSSSSDSDSVVKKPPGRKPAEKPLPKPRG 330
QY 104 NLPPKR--AKTEDEKEQRRVRLNRRRAAQSRRERKLEVEALEKRNKELETLLINVQ 161
Db 331 RKPKEPSPSSSSSDSDSDVDRISEWKR---RDEARRRELEARRRRQEBELRLRQ 386
QY 162 KTNLLIIVLELNRRSSGVVTRSS--SPLDL-SDSITLSQQLFGSRDQ-----TWS 211
Db 387 EKE--EKERRERADRGEAEAGSGSSGDELREDDEPVKKR---GRKGRGPPSSSDS 440
QY 212 NPEQSLMDQIMRSAANPTVNPASLSPISLPISDKFQTKEEDEBEQADEEMEQTWHETK 271
Db 441 EPEAELEERAKSAKPP--QSSSTEPARKP-GQKEKRVPEEKQQA-KPVKVERT-RKRS 495
QY 272 EAAAAKEKNSKQSRVSTD 289
Db 496 EGFSDMRKVEKKEPSVE 513

RESULT 3

US-10-131-826A-160
; Sequence 160, Application US/10131826A
; Publication No. US20050245730A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc

; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330RJ128
; CURRENT APPLICATION NUMBER: US/10/131,826A
; CURRENT FILING DATE: 2002-04-24
; PRIOR APPLICATION NUMBER: 60/049911
; PRIOR FILING DATE: 1997-06-18
; PRIOR APPLICATION NUMBER: 60/056974
; PRIOR FILING DATE: 1997-08-26
; PRIOR APPLICATION NUMBER: 60/059113
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059115
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059117
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059122
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059184
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059263
; PRIOR FILING DATE: 1997-09-18
; PRIOR APPLICATION NUMBER: 60/059352
; PRIOR FILING DATE: 1997-09-19
; PRIOR APPLICATION NUMBER: 60/059588
; PRIOR FILING DATE: 1997-09-19
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 160
; LENGTH: 605
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-131-826A-160

Query Match 6.0%; Score 136.5; DB 1; Length 605;
Best Local Similarity 20.2%; Pred. No. 0.0073;
Matches 104; Conservative 62; Mismatches 171; Indels 177; Gaps 22;

QY 7 SPLVKFEASPAESFLSA-----PGD---NFTSLFADSTPSTLNPRDMWTPDSVADI-- 54
Db 15 SPLLLLLVGG--FLGACVAGDEFGPEGLTSTLDDLLPTGLEPLDSEEPSETWGLGA 72
QY 55 -----DSRLSVIP-----ESQDAEDDESHSTSATA-----PSTSEKKPVKKR 91
Db 73 GLGASGSGFPSENEESRILQPPQVFWEEBELNDSSLDLPTADYVFPDLTEKAGSIED 132
QY 92 KSWGQVLP-----EPTNLPPRKRAKTEDEKEQRRVRLNRRRAAQSRRERK 139
Db 133 TSQAQELPNLPSPLPKMNLVEPPHMPHPPREEEEEEEEER----- 173
QY 140 RLEVEALEKRNKELETLLINVQKTNLLIIVLELNRRSSGVVTRSSSPDLSQDSITLSQ 199
Db 174 --EKEVEKQEEEEEBELLPVNGSQ-----EAKPQVRDFSLTSSSQTP----- 215
QY 200 QLFGSRDGTMSNPQSLMDQIMRSAANPTVNPASLSPISLPISDKFQTKEEDEBEQADE 259
Db 216 -----GATKSRHEDSGQASSGVVESSMGPSLLLPVTPPT-----VTPGDQ 258
QY 260 DEEMEQTWHETKEAAAAAKEKNSKQSRVSTDSTORPAVSGGDAAVP-VFSDDAGANCLGL 318
Db 259 DSTSQE-----AETATVLPAAGLGVFEAPQEAASEATAGAAGL 296


```
QY 319 DPVHQDDGPPSIGHSGLSAALDADRYLLESQSLASPNASTVDDDDYLAGDSAACTNPL- 377
Db 297 SQHSEVPALP--SF-----POTAPSGAEHPDEDPGSRSTA--SSPLA 337
QY 378 PSYDYF-----DINFLTDANHAA-----VDIVAASNYAAAD-- 410
Db 338 PGDMELTPSATLGOEDLNQOLLEGQAAEQSRIPWDSTOVICKWNSNLAKGNYIILNMT 397
QY 411 REIDLRIHDPENQIPSRHSIQ---PQSGASSHG 441
Db 398 ENIDCEVF--ROHRGPOLLALVEEVLPRHSGHHG 430

RESULT 4
US-10-485-517-129
; Sequence 129, Application US/10485517
; Publication No. US20050256299A1
; GENERAL INFORMATION:
; APPLICANT: University of Sheffield
; APPLICANT: Biosynexus Incorporated
; APPLICANT: Foster, Simon
; APPLICANT: Mond, James
; TITLE OF INVENTION: Antigenic Polypeptides
; FILE REFERENCE: P100629WQ
; CURRENT APPLICATION NUMBER: US/10/485,517
; CURRENT FILING DATE: 2004-02-02
; PRIOR APPLICATION NUMBER: GB 0118825.9
; PRIOR FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: GB 0200349.9
; PRIOR FILING DATE: 2002-01-09
; NUMBER OF SEQ ID NOS: 424
; SOFTWARE: Patent in version 3.1
; SEQ ID NO 129
; LENGTH: 895
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-10-485-517-129

Query Match 5.4%; Score 124.5; DB 1; Length 895;
Best Local Similarity 22.2%; Pred. No. 0.076;
Matches 98; Conservative 63; Mismatches 175; Indels 105; Gaps 21;

QY 55 DSRSLVIPESQDAEDDESHTSATSTSEKKPKVKRKGWGO-----VLPEPKT 103
Db 405 DKKLPELVSYDSKDYAIRFPVPSGTREVKIVSSIE-YGENIHBDYDYLTVFAQPIT 463
QY 104 NLP-----PRKAKTEDEK--BORRVERLNRRAAQSSRRERKLEVEA 145
Db 464 NNPDDVDEETYNLQKLLAPYHKAKTLERQVYLEKLEKLPKYKAEYKKKLDQTRVE- 522
QY 146 LEKRKKELETLINVKQTNLILVEELNRRFRSSGVVTRSSPLDSLQDSITLSQQLFGSR 205
Db 523 LADQVKSATVEFENVPTN---DQLTDLQEAHFVVFESEENSESVMDGFVEHPFYATL 578
QY 206 DQO---TMNPEQSL-MQIMRSAANPTV--NPASLSPSL--PPISDK----- 245
Db 579 NGQYVVMKTKDSDYKWLIVGKRVTVTSKDPKNSRTLIFPIYIPDKAVYNAIVKVVVA 638
QY 246 -----EFQKKEDEEQADEEME-QTWHETKEAATAKKNQKQSRVSTD 289
Db 639 NIGYEQYHVRINQDINTKDDTSQNTSEPLNVGTQGEKGVADTVVAENSGSTATNPKD 698
QY 290 STORPAVSIIGDAAVPFSD---DAGANCLGLDPVHVHQDDGPPSIGHSGLSAALDADRYL 346
Db 699 ASDKADV-----IEPESDVVKDADNN---IDKDVQHD-----VDHLSDMSDNNHFDKYD 744
QY 347 LESQ---LLASPNASTVDDDDYLAGDSAACTNPLPSDYDFDIN-----DFLTDANH-- 395
Db 745 LKEMDTQIAKDTDRNYDKO---ADNSVGMSSNVDTDKDSNKNKDKVQLNHIADKNHGTG 801
QY 396 --AAYDIVAASNYAAADRELD 414
```

```
Db 802 KKAALDVV-KQNYNNTDKYTD 821

RESULT 5
US-10-821-234-1120
; Sequence 1120, Application US/10821234
; Publication No. US20050255114A1
; GENERAL INFORMATION:
; APPLICANT: Labat, Ivan
; APPLICANT: Stache-Crain, Birgit
; APPLICANT: Andarmani, Susan
; APPLICANT: Tang, Y. Tom
; TITLE OF INVENTION: Methods for Diagnosis and Treatment of Preeclampsia
; FILE REFERENCE: 821A
; CURRENT APPLICATION NUMBER: US/10/821,234
; CURRENT FILING DATE: 2004-04-07
; PRIOR APPLICATION NUMBER: US 60/462,047
; PRIOR FILING DATE: 2003-04-07
; NUMBER OF SEQ ID NOS: 1704
; SOFTWARE: pt_seq_genes Version 1.0
; SEQ ID NO 1120
; LENGTH: 4384
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-821-234-1120

Query Match 5.4%; Score 123.5; DB 1; Length 4384;
Best Local Similarity 22.2%; Pred. No. 0.61;
Matches 100; Conservative 58; Mismatches 202; Indels 91; Gaps 22;

QY 53 DIDSRLVIPESQD-AEDDESHTSATSTSEKKPKVKRKGWQVLPEPKTNLPKRKA 111
Db 2996 ELSQKLSSQSMKSTVETQHFNSIEDKVYSEISKVSHQSYVGL-----CPPLEET 3048
QY 112 KTEDEKEQRVERVLNRRAAQSRRERKLEVEALEK-----RNKELETLLINVKQT 163
Db 3049 ETSPTKSPDSLE--FSPGKESPSDDVFDHSPIDGLEKLAPLAQTEGKEIKTLPVYVSFV 3106
QY 164 NL-----ILVEELNRRFRSSGV--VTRSS-----SPLDSLQDSITLSQQLF 202
Db 3107 QVGKQVEKEIQGGVKKIIISQBECKTVQETRGTPYTTROQKQPPSQSPGSD--TLEQVSF 3165
QY 203 GSRDQOTNSPEQSLMDQIMRSAANPTVNPASLSPSLP--PISDKFQTKKEDEEQADE 260
Db 3166 LBSGKSPKPLTPETPSEEVSEVYFTSKT--PDSLIAVIPKSPDIPVESESEEEQA-KS 3222
QY 261 EMEQOTWHETKEAATAKKNQKQSRVSTDSTORPAVSIIGDAAVPFVSDDAGANCLGLDP 320
Db 3223 TSLKQT--TVEETAVERE---MPNDVSKDSNQRPK---NNRVAYIEFPFPPPP---LOADQ 3271
QY 321 VHODGPPSIGHSGLSAALDA-----DRYLESQSLASPNASTVDDDDYLAGDSA- 370
Db 3272 IESDK-----KHHYLPKEKVDMEIVNLQDEHDKYQLAEFVIRVQPPSPVPPGADVSDSD 3326
QY 371 ACFTNPLP-SDYDFDINDFLTDD-----ANHAAYDIVAASNYAAADRELDLEIHD 419
Db 3327 DESIYQVPVVKYTKFLKE--VDDEQEKPKASAEKASNQKELESNGSKDNFGLGLDS 3384
QY 420 PENQIPSRHSIQPOSGAS-----SHGCD 443
Db 3385 PQNEIAQNGNDQSITECSIATTAEFSDHTD 3415

RESULT 6
US-10-485-517-212
; Sequence 212, Application US/10485517
; Publication No. US20050256299A1
; GENERAL INFORMATION:
; APPLICANT: University of Sheffield
; APPLICANT: Biosynexus Incorporated
; APPLICANT: Foster, Simon
; APPLICANT: Mond, James
; TITLE OF INVENTION: Antigenic Polypeptides
```

```

; FILE REFERENCE: P100623WO
; CURRENT APPLICATION NUMBER: US/10/485,517
; CURRENT FILING DATE: 2004-02-02
; PRIOR APPLICATION NUMBER: GB 0118825.9
; PRIOR FILING DATE: 2001-08-02
; PRIOR APPLICATION NUMBER: GB 0200349.9
; PRIOR FILING DATE: 2002-01-09
; NUMBER OF SEQ ID NOS: 424
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 212
; LENGTH: 1448
; TYPE: PRP
; ORGANISM: Staphylococcus aureus
US-10-485-517-212

Query Match          5.4%; Score 123; DB 1; Length 1448;
Best Local Similarity 19.7%; Pred. No. 0.17; Indels 78; Gaps 10;
Matches 67; Conservative 51; Mismatches 144;

QY   33 ADSTPTSLNPRDMTPDSVA-----DIDRLSV-- 60
      ||:||:||:||:||:||:||:||:||:||:||:||:||:||:||:||:|||||
Db   1079 ADASPTPNNSDAATGTTATSATDANDKPAQNSSVDASTNPTMNDVTSKPEVES 1138

QY   61 -----IPSQDAEDDES-HSTSATAPSTSEKKPKVKRKSQGVLPEPKTNLPPKR 110
      :||::||::||::||::||::||::||::||::||::||::||::||::||
Db   1139 TNGTTPDKVTEDNATPAESTNNSTTTATNENAPT-----GSTATAPTT--ASTE 1190

QY   111 AKTEDEKEQRVERVLNRRAAQSRERKRLEVEALEKRKNKELETLLINVOKTNLII 170
      :||::||::||::||::||::||::||::||::||::||::||::||::||
Db   1191 ASSADSKMASVNDKQNAEVNNSAESQSTNDKVAQPKSENKAKEKDGSNSTNQSWES 1250

QY   171 LNRFRSSGVWTRSSSPDLSDLSQTLFSGRDGTMSNPQSMDQT---MRSAA 226
      ||:||:||:||:||:||:||:||:||:||:||:||:||:||:||:||:|||||
Db   1251 TTETLPSAD-ITFPVPNTSKDK-----BESITTQTDAGLKSET 1290

QY   227 NPTVNPASLPSPPLISDFEFTKBEDEQEAEEMEOTWHETKEAATAKERKNSOSRV 286
      ||:||:||:||:||:||:||:||:||:||:||:||:||:||:||:||:|||||
Db   1291 NVASNEADKPSK--ADTEVSNKPSTSSASSEAKEKMTSTNVSOQDDTATADTNDTKSV 1347

QY   287 STDSTORPAVSTGGDAANPVFSDDA-GANCLGLDPVHODD 325
      ||:||:||:||:||:||:||:||:||:||:||:||:||:||:||:||:|||||
Db   1348 GSAANNKATONDGANASPATVNSGNNSANQDMLNVTNTDD 1387

RESULT 7
US-10-821-234-1285
; Sequence 1285, Application US/10821234
; Publication No. US20050255114A1
; GENERAL INFORMATION:
; APPLICANT: Labat, Ivan
; APPLICANT: Stache-Crain, Birgit
; APPLICANT: Andarmani, Susan
; APPLICANT: Tang, Y. Tom
; TITLE OF INVENTION: Methods for Diagnosis and Treatment of Preclampsia
; FILE REFERENCE: 821A
; CURRENT APPLICATION NUMBER: US/10/821,234
; CURRENT FILING DATE: 2004-04-07
; PRIOR APPLICATION NUMBER: US 60/462,047
; PRIOR FILING DATE: 2003-04-07
; NUMBER OF SEQ ID NOS: 1704
; SOFTWARE: pt seq_genes Version 1.0
; SEQ ID NO 1285
; LENGTH: 431
; TYPE: PRP
; ORGANISM: Homo sapiens
US-10-821-234-1285

Query Match          5.2%; Score 119.5; DB 1; Length 431;
Best Local Similarity 23.1%; Pred. No. 0.068; Indels 61; Gaps 13;
Matches 64; Conservative 45; Mismatches 107;

QY   14 AGPAESFLS---APGDNFTSLPADSTPTSLNPRDMTPDSVADID-----SRLSVIPESQ 65
      |||||

```



```
QY 11 KFEASPAESFLSAPGDNFTSLFADSTPST-LNPRDMTTPDSVADIDSRLSVIP-ESQDAE 68
Db 12 RMESPLSRAPSRGGVNFNLNARTYIPNTKVECHYTLPPGTMPASDWDGIFKVEACVR 71
QY 69 DDESHSTSATPSTSEKKVPKRSW-GQVLPPEKTNL-----PPKRA 111
Db 72 DYHTFWSSVPSTTSGSPIHTSVQFQASVLPKPGAQLYQFRVYNRQGVCGSQPPQFR 131
QY 112 KTEDEKEORRVER-----VLNRRAAQSRRERKRL-----EVALE 147
Db 132 EPRPMDLVTLLEADGGSDILLVVPKATVLQN-QLDESOQERNLMLQLQLEGQVTEL 190
QY 148 KRNELETLLINVOKTNLLILVELNFRSSGVVTRSSPLD-----SLQDSI- 195
Db 191 SRVQELERALARATQAEHTLMEOYKIGISRGEITEERDILSRQGDHVARILEDDIQ 250
QY 196 TLSQOLFSGS-----RDG-QTMSNPEQSLMDQIMRSAAANTVNPASISPLSPISKEF 247
Db 251 TISEKVLTKVELDLRLDTVKALTRQEKLGLQKVEQADKEQ-----SEAL 298
QY 248 QTKEDEQAQDEDEMEQTHETKEAAAKEKN-----SKQSRVS----- 287
Db 299 QVAQENHHLNLDLKEAKSWOEQSAQARLKDQVAKMDTLGQAQORVAEPLKEQLR 358
QY 288 -----TDSTORPAVSIIGDAA 303
Db 359 GAQLAASSQKATLIGBELA 379
```

```
RESULT 12
US-10-131-826A-16
; Sequence 16, Application US/10131826A
; Publication No. US20050245730A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tamas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330R1C128
; CURRENT APPLICATION NUMBER: US/10/131, 826A
; CURRENT FILING DATE: 2002-04-24
; PRIOR APPLICATION NUMBER: 60/049911
; PRIOR FILING DATE: 1997-06-18
; PRIOR APPLICATION NUMBER: 60/056974
; PRIOR FILING DATE: 1997-08-26
; PRIOR APPLICATION NUMBER: 60/059113
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059115
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059117
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059122
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059184
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059263
; PRIOR FILING DATE: 1997-09-18
```

```
; PRIOR APPLICATION NUMBER: 60/059352
; PRIOR FILING DATE: 1997-09-19
; PRIOR APPLICATION NUMBER: 60/059588
; PRIOR FILING DATE: 1997-09-19
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 16
; LENGTH: 691
; TYPE: PRT
; ORGANISM: Homo sapien
US-10-131-826A-16

Query Match 4.9%; Score 111.5; DB 1; Length 691;
Best Local Similarity 19.0%; Pred. No. 0.42;
Matches 72; Conservative 61; Mismatches 145; Indels 101; Gaps 14;

QY 13 EASPAESFLSAPGDNFTSLFADSTPST-LNPRDMTTPDSVADIDSRLSVIP-ESQDAEDD 70
Db 2 EESPLSRAPSRGGVNFNLNARTYIPNTKVECHYTLPPGTMPASDWDGIFKVEACVRDY 61
QY 71 ESHSTSATPSTSEKKVPKRSW-GQVLPPEKTNL-----PPKRAKT 113
Db 62 HTFWSSVPSTTSGSPIHTSVQFQASVLPKPGAQLYQFRVYNRQGVCGSQPPQPREP 121
QY 114 EDEKEORRVER-----VLNRRAAQSRRERKRL-----EVALEKR 149
Db 122 RPDMLVTLLEADGGSDILLVVPKATVLQN-QLDESOQERNLMLQLQLEGQVTELRSR 180
QY 150 NKELETLLINVOKTNLLILVELNFRSSGVVTRSSPLD-----SLQDSI-TL 197
Db 181 VQELERALARATQAEHTLMEOYKIGISRGEITEERDILSRQGDHVARILEDDIQTI 240
QY 198 SQOLFSGS-----RDG-QTMSNPEQSLMDQIMRSAAANTVNPASISPLSPISKEFOT 249
Db 241 SEKVLTKVELDLRLDTVKALTRQEKLGLQKVEQADKEQ-----SEALEQV 288
QY 250 KEEDEQAQDEDEMEQTHETKEAAAKEKN-----SKQSRVS----- 287
Db 289 AQENHHLNLDLKEAKSWOEQSAQARLKDQVAKMDTLGQAQORVAEPLKEQLRGA 348
QY 288 ---TDSTORPAVSIIGDAA 303
Db 349 QELAASSQKATLIGBELA 367
```

```
RESULT 13
US-10-821-234-1477
; Sequence 1477, Application US/10821234
; Publication No. US20050255114A1
; GENERAL INFORMATION:
; APPLICANT: Labat, Ivan
; APPLICANT: Stache-Crain, Birgit
; APPLICANT: Andarmani, Susan
; APPLICANT: Tang, Y. Tom
; TITLE OF INVENTION: Methods for Diagnosis and Treatment of Preeclampsia
; FILE REFERENCE: 821A
; CURRENT APPLICATION NUMBER: US/10/821,234
; CURRENT FILING DATE: 2004-04-07
; PRIOR APPLICATION NUMBER: US 60/462,047
; PRIOR FILING DATE: 2003-04-07
; NUMBER OF SEQ ID NOS: 1704
; SOFTWARE: pt seq_genes Version 1.0
; SEQ ID NO 1477
; LENGTH: 667
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-821-234-1477
```

```
Query Match 4.8%; Score 111; DB 1; Length 667;
Best Local Similarity 21.0%; Pred. No. 0.43;
Matches 101; Conservative 57; Mismatches 168; Indels 154; Gaps 21;

QY 5 QSSPLVKFEASPAESFLSAPGDNFTSLFADSTPSTLNPRDMTTPDSVADIDSRLSVIPES 64
```

```
Db 222 QAGTVRSAPPOEKKRVEKDTMSQDALEALSGLTRQ-----APEFLDLSIKEV 274
QY 65 QDA-----EDESHSTSATAPSTSEKPKVKRKSQVLPPEPKTNLPKRKRK- 112
Db 275 DEAKAKEEKLKGEDE-----TIPSYRLKPATD-KDGKPLLPPEPEKPKRSESEL 327
QY 113 ----TED-----EKEQRRVERVLNRRRAQS--SRERKRLEVEALEKRNKELETLLIN 159
Db 328 IDLSDFDRSCKEKPSTKTEKESKAAAPAPVSEAVCRTSMCSIQSAPPEPATL--- 384
QY 160 VQKTNLILVEELNRRFRSSGVVTRSSPLDSIQDSITLSQQLFGSRDGTMSNPEQSLMD 219
Db 385 -----KGTV--PDDAVEALADSL-----GKKEADPEDG--- 410
QY 220 QIMRSAANTVNPASLSPSLPISDK-EFQTKEEDEEQADEDEMEQWTHETKEAAAAKE 278
Db 411 -----KPVMDKVKKEKEDEKLEKEETIPPDYRLKEE---VKD 447
QY 279 KNSKQSRVSTSTORPAYSIGG--DAAVPVFSDDAGANCLGLDPVHQDDGPFSGHSGFL 336
Db 448 KDGKPLLPKESKEQPPMSDEFLDALSDFGSQPNASSLKFEADK-----L 494
QY 337 SAALDADRYLLESQLLASPNASTVDDDYLAGDSAACTFNPLPSDYDF-DINDFLTD---- 391
Db 495 AAAI-----SEVVSQTPASTTQAGAPRPTS-----QSDKDLDDLKLSLQ 539
QY 392 -----DANHAAYDIVAASNAADRELDEIHDPENQIPSRHSIQOQSGASSHCGDDGG 446
Db 540 RQDPDPENKPMEDKVKKEK-----AKAHRDKLGERDDTIPPEY-----RHLLDDNG 585

RESULT 14
US-10-821-234-1514
; Sequence 1514, Application US/10821234
; Publication No. US20050255114A1
; GENERAL INFORMATION:
; APPLICANT: Labat, Ivan
; APPLICANT: Stache-Crain, Birgit
; APPLICANT: Andarmani, Susan
; APPLICANT: Tang, Y. Tom
; TITLE OF INVENTION: Methods for Diagnosis and Treatment of Preeclampsia
; FILE REFERENCE: 821A
; CURRENT APPLICATION NUMBER: US/10/821,234
; CURRENT FILING DATE: 2004-04-07
; PRIOR APPLICATION NUMBER: US 60/462,047
; PRIOR FILING DATE: 2003-04-07
; NUMBER OF SEQ ID NOS: 1704
; SOFTWARE: pt_seq_genes Version 1.0
; SEQ ID NO 1514
; LENGTH: 915
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-821-234-1514

Query Match 4.8%; Score 111; DB 1; Length 915;
Best Local Similarity 22.6%; Pred. No. 0.63;
Matches 82; Conservative 52; Mismatches 145; Indels 84; Gaps 15;

QY 29 TSLPADSTPSTLN-PRDMWTPDSVADIDSRLSVIPESQ-----DABDDSHSTSATAP 80
Db 501 TSSGKSTITRVNSPGLTARLGSVTHVTSFSHAPSSRGGCSIKWEAPPAEPLAAVEAA 560
QY 81 STSEKKPKVKRKSQVLPPEKTNLPKRKRKTEDEK-----EQRRVERV--- 125
Db 561 NGAQTRVNVK-----APEGRSPLSABELMTIEDEGLDKMLDQSTDFEERKLIRAAALR 613
QY 126 -LRNPRRAQSRERKRLEVEALEKRNKELETLLINQKTNLILVEELNRRFRSSGVVTR 184
Db 614 ELRQKRQDQRDKERERLQEA---RGRFGE-----GRGNTATETTTTR 653
QY 185 SSPL--DSLQDSITLSQQLFGSRDGTMSNPEQSLMDQIMRSAAN-----PTVNP 233
```

Search completed: November 23, 2005, 03:32:37
Job time : 7.71544 secs

```
Db 654 SORAAADGSAVSTVTYTERLHVHNDG-TRTARTTTVSSFVRSENGSGSTMOTKTFSSS 712
QY 234 SLSPSLPISDKPEFQTKEE-----DEEQADEDEM--EQTWTHETKEAAAAK---EKNS 281
Db 713 SSSKMGSIFFREDQASPRAGSLAALEKRAQAEKKELMKAQSLPKTASQARKAMIEKLE 772
QY 282 KQSRVSTSTORPAYSIGGDAVP-----VPSDDAGANCLGLDPVHQDDGPFSGHSGP 334
Db 773 KEGAAGSGGPRAAVQVSTSGFVNPANSIKOMLLDWCRAKTRGYE--HVDIQNFSSSWSD 830
QY 335 GLS 337
Db 831 GMA 833

RESULT 15
US-10-793-626-1780
; Sequence 1780, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU3480US
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1780
; LENGTH: 1155
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: amino acid sequence
US-10-793-626-1780

Query Match 4.8%; Score 110; DB 1; Length 1155;
Best Local Similarity 23.2%; Pred. No. 0.98;
Matches 92; Conservative 43; Mismatches 190; Indels 72; Gaps 14;

QY 63 ESQDAEDDESHTSATAPSTSEKPKVKRKS-----GQVLPKPTNLPKRKAK----- 112
Db 49 EAKAEDNQLSESASKEEQGSRDNESSKLNQVLDNGSHSEKTTNVNNATEVKVEAPT 108
QY 113 TEDEKEQRRVERVLNRRRAQSSRERKRLEVEALEKRNKELETLLINQKTNLILVEELN 172
Db 109 TSDVSKPKANEAVVNESTKPKTTEAPTVEESIAETPKTSTTQODSTEKKNPSLKDNLN 168
QY 173 RFRSSGVVTRSSPLDSLQDSITLSQQLFGSRDGTMSNPEQSLMDQIMRSAAN--PTV 230
Db 169 -----SSSTTSKESKTD--EHSTKQAQMSNTKNSLNDTNDSPQTS---EKTSSQANNDSTD 218
QY 231 NPASLSGL--PPISDKPEFQTKEEDEEQADEDEMEQWTHETKEAAAAKEKNSKQSRVST 288
Db 219 NQSAQSKQLDSKPSQKPKYKTKFNDE--PTQDQVHTHTKLKTPSISTDSSVNDKQ----- 271
QY 289 DSTORPAYSIGGDAVPVPSDDAGANCLGL-----DPAHQDDG 326
Db 272 DYTSASVASLGVDSNETEATINAVRDNLDLKAASREQINEAIIAEALKKDFSNPDYVDVT 331
QY 327 PFSIGHSGFL--SAALDADRYLLESQLLASPNASTVDDDYLAGDSAACTNPLPSDYDFDI 385
Db 332 PLALNTSQSKNSPHKSASPRMNLMSLAAEPNSGKNVNDKVK-----ITNPTLS----- 379
QY 386 NDFLTDDANHAAYDIVAASN---YAAADRELDEIHD 419
Db 380 ---LNKSNHANNHWIPTSNEQFNLKANYELDDSDIKE 413
```

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

QM protein - protein search, using sw model

Run on: November 23, 2005, 03:12:56 ; Search time 146.003 Seconds
(without alignments)
1290.661 Million cell updates/sec

Title: US-10-663-450-2

Perfect score: 2294

Sequence: 1 MAFQSSPLVKFEASPAESF.....QPQSGASSHCDDGGIAGV 451

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA.Main:

- 1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pap:*
- 2: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pap:*
- 3: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pap:*
- 4: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pap:*
- 5: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pap:*
- 6: /cgn2_6/ptodata/1/pubpaa/US11_PUBCOMB.pap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2294	100.0	451	3	US-09-816-277-2
2	2294	100.0	451	3	US-09-816-277-5
3	2294	100.0	451	4	US-10-663-450-2
4	2294	100.0	451	4	US-10-663-450-5
5	434.5	18.9	386	3	US-09-816-277-16
6	434.5	18.9	386	4	US-10-663-450-16
7	434	18.9	349	3	US-09-816-277-4
8	434	18.9	349	3	US-09-816-277-6
9	434	18.9	349	4	US-10-663-450-4
10	434	18.9	349	4	US-10-663-450-6
11	429.5	18.7	409	3	US-09-533-029-104
12	429.5	18.7	409	4	US-10-295-403-158
13	429.5	18.7	409	4	US-10-412-699B-684
14	426.5	18.6	342	3	US-09-816-277-19
15	426.5	18.6	342	4	US-10-663-450-19
16	312.5	13.6	174	4	US-10-767-701-61238
17	312.5	10.1	200	4	US-10-369-493-21867
18	196	8.5	68	3	US-09-816-277-60
19	196	8.5	68	4	US-10-663-450-60
20	168	7.2	3257	6	US-11-097-143-29398
21	165.5	7.2	132	3	US-09-772-656-6
22	165.5	7.2	132	3	US-09-772-656-10
23	165.5	7.2	192	4	US-10-396-199A-6
24	165.5	7.2	192	4	US-10-396-199A-10
25	163.5	7.1	1442	5	US-10-450-763-55960
26	163.5	7.1	2408	5	US-10-450-763-40990
27	163	7.1	132	3	US-09-772-656-2

28	163	7.1	192	4	US-10-396-199A-2	Sequence 2, Appli
29	148.5	6.5	650	3	US-09-801-368-430	Sequence 430, App
30	148.5	6.5	2048	3	US-09-815-379-12	Sequence 12, Appl
31	148.5	6.5	2057	3	US-09-815-379-10	Sequence 10, Appl
32	147.5	6.4	2058	3	US-09-815-379-17	Sequence 17, Appl
33	147.5	6.4	2058	4	US-10-021-660-120	Sequence 120, Appl
34	147.5	6.4	2058	4	US-10-211-462-207	Sequence 207, App
35	147.5	6.4	2058	4	US-10-648-593-188	Sequence 188, App
36	147.5	6.4	2058	5	US-10-723-860-742	Sequence 742, App
37	146.5	6.4	1017	4	US-10-032-585-7590	Sequence 7590, Ap
38	145.5	6.3	639	4	US-10-697-787-14	Sequence 14, Appl
39	145.5	6.3	639	5	US-10-512-600-14	Sequence 14, Appl
40	145.5	6.3	5333	6	US-11-097-143-24108	Sequence 24108, A
41	145.5	6.3	5560	4	US-10-263-929-142	Sequence 142, App
42	145.5	6.3	5560	6	US-11-097-143-40272	Sequence 40272, A
43	145	6.3	1000	4	US-10-128-714-3305	Sequence 3305, Ap
44	144.5	6.3	665	4	US-10-425-115-190929	Sequence 190929,
45	144.5	6.3	672	4	US-10-425-114-58469	Sequence 58469, A

ALIGNMENTS

RESULT 1

US-09-816-277-2

; Sequence 2, Application US/09816277

; Publication No. US20010034045A1

; GENERAL INFORMATION:

; APPLICANT: Penttila, Merja E.

; APPLICANT: Ward, Michael

; APPLICANT: Wang, Huaming

; APPLICANT: Valkonen, Mari J.

; APPLICANT: Saloheimo, Markku

; TITLE OF INVENTION: Increased Production of Secreted

; FILE OF INVENTION: Proteins by Recombinant Eukaryotic Cells

; FILE REFERENCE: GC590-2

; CURRENT APPLICATION NUMBER: US/09/816,277

; PRIOR FILING DATE: 2001-03-23

; PRIOR APPLICATION NUMBER: US 09/534,692

; PRIOR FILING DATE: 2000-03-24

; NUMBER OF SEQ ID NOS: 63

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 2

; LENGTH: 451

; TYPE: PRT

; ORGANISM: Trichoderma reesei

; US-09-816-277-2

Query Match 100.0%; Score 2294; DB 3; Length 451;

Best Local Similarity 100.0%; Pred. No. 2.6e-143; Mismatches 0; Indels 0; Gaps 0;

Matches 451; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAFQSSPLVKFEASPAESFSLGAPGNFTSLFADSTPSTLNPRDMMTPDSVADIDRLSV 60

Db 1 MAFQSSPLVKFEASPAESFSLGAPGNFTSLFADSTPSTLNPRDMMTPDSVADIDRLSV 60

QY 61 IPESQDAEDDESHSTSATAPSTSEKPKVKRKSQGVLPPEKTNLPKRKAETDEKEQR 120

Db 61 IPESQDAEDDESHSTSATAPSTSEKPKVKRKSQGVLPPEKTNLPKRKAETDEKEQR 120

QY 121 RVERVLNRRAAQSSRRERKLEVEALEKRNKELETLLINVQKTNLLILVEELNFRSSGV 180

Db 121 RVERVLNRRAAQSSRRERKLEVEALEKRNKELETLLINVQKTNLLILVEELNFRSSGV 180

QY 181 VTRSSPLDLSODITLSQOLFGRDQGTWMSNPESQSLMDQIMRSAANPTVNPASLSPSLP 240

Db 181 VTRSSPLDLSODITLSQOLFGRDQGTWMSNPESQSLMDQIMRSAANPTVNPASLSPSLP 240

QY 241 PISDKREFQTKBEDEQADEMEQTHETKEAAAAAKENSKOSRVSSTQTPAVSIGG 300

Db 241 PISDKREFQTKBEDEQADEMEQTHETKEAAAAAKENSKOSRVSSTQTPAVSIGG 300

QY 301 DAAVPFSDAGANCLGLDPVHQDDGPFSGHSGFGLSALDADRYLLESQALLASPNASTV 360

Db 301 DAAVPVFSDDAGANCLGLDPVHQDDGPFSGHSGFGLSAALDADRYLLESQLLASPNASTV 360
QY 361 DDYLAGDSAACTNPLPSDYDFDINDFLTDDANHAAYDIVAASNYAAADRELDLEIHP 420
Db 361 DDYLAGDSAACTNPLPSDYDFDINDFLTDDANHAAYDIVAASNYAAADRELDLEIHP 420
QY 421 ENQIPSRHSIIQQPSGASSHGCDGGIAVGV 451
Db 421 ENQIPSRHSIIQQPSGASSHGCDGGIAVGV 451

RESULT 2

US-09-816-277-5
; Sequence 5, Application US/09816277
; Publication No. US20010034045A1
; GENERAL INFORMATION:
; APPLICANT: Penttila, Merja E.
; APPLICANT: Ward, Michael
; APPLICANT: Wang, Huaming
; APPLICANT: Valkonen, Mari J.
; APPLICANT: Saloheimo, Markku
; TITLE OF INVENTION: Increased Production of Secreted
; TITLE OF INVENTION: Proteins by Recombinant Eukaryotic Cells
; FILE REFERENCE: GC590-2
; CURRENT APPLICATION NUMBER: US/09/816,277
; CURRENT FILING DATE: 2001-03-23
; PRIOR APPLICATION NUMBER: US 09/534,692
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 451
; TYPE: PRT
; ORGANISM: Trichoderma reesei
US-09-816-277-5

Query Match 100.0%; Score 2294; DB 3; Length 451;
Best Local Similarity 100.0%; Pred. No. 2.6e-143;
Matches 451; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MAFQSSPLVKFEASPAESFLSAPGDNFTSLPADSTPSTLNPRDMTPSPSVADIDSRLSV 60
Db 1 MAFQSSPLVKFEASPAESFLSAPGDNFTSLPADSTPSTLNPRDMTPSPSVADIDSRLSV 60
QY 61 IPESQDAEDDESHSTSATAPSTSEKKPVKKRSGQVLPKPTNLPKRKTEDEKEOR 120
Db 61 IPESQDAEDDESHSTSATAPSTSEKKPVKKRSGQVLPKPTNLPKRKTEDEKEOR 120
QY 121 RVERVLNRRAAQSSRKRLEVEALEKRNKELETLLINVQKTNLILVEELNFRFRSSGV 180
Db 121 RVERVLNRRAAQSSRKRLEVEALEKRNKELETLLINVQKTNLILVEELNFRFRSSGV 180
QY 181 VTRSSPLDSLQDSITLSQQLFGSRDGTMSNPEQSLMDQIMRSAANPTVNPASLSPSLP 240
Db 181 VTRSSPLDSLQDSITLSQQLFGSRDGTMSNPEQSLMDQIMRSAANPTVNPASLSPSLP 240
QY 241 PISDKFOTKEDEEQADEDEMEQTHETKEAAAKEKNKQSRVSTSTORPAVSIGG 300
Db 241 PISDKFOTKEDEEQADEDEMEQTHETKEAAAKEKNKQSRVSTSTORPAVSIGG 300
QY 301 DAAVPVFSDDAGANCLGLDPVHQDDGPFSGHSGFGLSAALDADRYLLESQLLASPNASTV 360
Db 301 DAAVPVFSDDAGANCLGLDPVHQDDGPFSGHSGFGLSAALDADRYLLESQLLASPNASTV 360
QY 361 DDYLAGDSAACTNPLPSDYDFDINDFLTDDANHAAYDIVAASNYAAADRELDLEIHP 420
Db 361 DDYLAGDSAACTNPLPSDYDFDINDFLTDDANHAAYDIVAASNYAAADRELDLEIHP 420
QY 421 ENQIPSRHSIIQQPSGASSHGCDGGIAVGV 451
Db 421 ENQIPSRHSIIQQPSGASSHGCDGGIAVGV 451

RESULT 3

US-10-663-450-2
; Sequence 2, Application US/10663450
; Publication No. US20040186070A1
; GENERAL INFORMATION:
; APPLICANT: Penttila, Merja E.
; APPLICANT: Ward, Michael
; APPLICANT: Wang, Huaming
; APPLICANT: Valkonen, Mari J.
; APPLICANT: Saloheimo, Markku
; TITLE OF INVENTION: Increased Production of Secreted
; TITLE OF INVENTION: Proteins by Recombinant Eukaryotic Cells
; FILE REFERENCE: GC590-2
; CURRENT APPLICATION NUMBER: US/10/663,450
; CURRENT FILING DATE: 2003-09-15
; PRIOR APPLICATION NUMBER: US 09/534,692
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 451
; TYPE: PRT
; ORGANISM: Trichoderma reesei
US-10-663-450-2

Query Match 100.0%; Score 2294; DB 4; Length 451;
Best Local Similarity 100.0%; Pred. No. 2.6e-143;
Matches 451; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MAFQSSPLVKFEASPAESFLSAPGDNFTSLPADSTPSTLNPRDMTPSPSVADIDSRLSV 60
Db 1 MAFQSSPLVKFEASPAESFLSAPGDNFTSLPADSTPSTLNPRDMTPSPSVADIDSRLSV 60
QY 61 IPESQDAEDDESHSTSATAPSTSEKKPVKKRSGQVLPKPTNLPKRKTEDEKEOR 120
Db 61 IPESQDAEDDESHSTSATAPSTSEKKPVKKRSGQVLPKPTNLPKRKTEDEKEOR 120
QY 121 RVERVLNRRAAQSSRKRLEVEALEKRNKELETLLINVQKTNLILVEELNFRFRSSGV 180
Db 121 RVERVLNRRAAQSSRKRLEVEALEKRNKELETLLINVQKTNLILVEELNFRFRSSGV 180
QY 181 VTRSSPLDSLQDSITLSQQLFGSRDGTMSNPEQSLMDQIMRSAANPTVNPASLSPSLP 240
Db 181 VTRSSPLDSLQDSITLSQQLFGSRDGTMSNPEQSLMDQIMRSAANPTVNPASLSPSLP 240
QY 241 PISDKFOTKEDEEQADEDEMEQTHETKEAAAKEKNKQSRVSTSTORPAVSIGG 300
Db 241 PISDKFOTKEDEEQADEDEMEQTHETKEAAAKEKNKQSRVSTSTORPAVSIGG 300
QY 301 DAAVPVFSDDAGANCLGLDPVHQDDGPFSGHSGFGLSAALDADRYLLESQLLASPNASTV 360
Db 301 DAAVPVFSDDAGANCLGLDPVHQDDGPFSGHSGFGLSAALDADRYLLESQLLASPNASTV 360
QY 361 DDYLAGDSAACTNPLPSDYDFDINDFLTDDANHAAYDIVAASNYAAADRELDLEIHP 420
Db 361 DDYLAGDSAACTNPLPSDYDFDINDFLTDDANHAAYDIVAASNYAAADRELDLEIHP 420
QY 421 ENQIPSRHSIIQQPSGASSHGCDGGIAVGV 451
Db 421 ENQIPSRHSIIQQPSGASSHGCDGGIAVGV 451

RESULT 4

US-10-663-450-5
; Sequence 5, Application US/10663450
; Publication No. US20040186070A1
; GENERAL INFORMATION:
; APPLICANT: Penttila, Merja E.
; APPLICANT: Ward, Michael
; APPLICANT: Wang, Huaming
; APPLICANT: Valkonen, Mari J.
; APPLICANT: Saloheimo, Markku


```
; TITLE OF INVENTION: Increased Production of Secreted
; FILE REFERENCE: GC590-2
; CURRENT APPLICATION NUMBER: US/10/663,450
; PRIOR FILING DATE: 2003-09-15
; PRIOR APPLICATION NUMBER: US 09/534,692
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 451
; TYPE: PRT
; ORGANISM: Trichoderma reesei
US-10-663-450-5

Query Match      100.0%; Score 2294; DB 4; Length 451;
Best Local Similarity 100.0%; Pred. No. 2.6e-143;
Matches 451; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1  MAFQSSPLVKFPASPAESFLSAPGDNFTSLPADSTPSTLNPRDMTPDSVADIDSRLSV 60
Db 1  MAFQSSPLVKFPASPAESFLSAPGDNFTSLPADSTPSTLNPRDMTPDSVADIDSRLSV 60
QY 61  IPESQDAEDDESHTSATAPSTSEKKPVKKRKSQGVLPPEKTNLPKRKAKTEDEKEOR 120
Db 61  IPESQDAEDDESHTSATAPSTSEKKPVKKRKSQGVLPPEKTNLPKRKAKTEDEKEOR 120
QY 121  RVERVLNRRAQSSRRERKRLVEALEKRNKELETLINVOKTNLILVEELNFRFRSSGV 180
Db 121  RVERVLNRRAQSSRRERKRLVEALEKRNKELETLINVOKTNLILVEELNFRFRSSGV 180
QY 181  VTRSSSPLDSLQDSITLSQQLFGSRDQGTMSNPESQIMDSQIMRSAANPTVNPASLSPSLP 240
Db 181  VTRSSSPLDSLQDSITLSQQLFGSRDQGTMSNPESQIMDSQIMRSAANPTVNPASLSPSLP 240
QY 241  PISDKSFQTEKEDEQADEEMEQTWHETKEAAAKKSKOSRVSTSTORPAVSIGG 300
Db 241  PISDKSFQTEKEDEQADEEMEQTWHETKEAAAKKSKOSRVSTSTORPAVSIGG 300
QY 301  DAAVPVFSDDAGANCLGLDPVHDDGPFSGHSGFLSAALDADRYLLESQLLASPNASTV 360
Db 301  DAAVPVFSDDAGANCLGLDPVHDDGPFSGHSGFLSAALDADRYLLESQLLASPNASTV 360
QY 361  DDYLAGDSAACTNPLPSDYDFINDFLTDDANHAAYDIVAASNVAADRELDLEIHD 420
Db 361  DDYLAGDSAACTNPLPSDYDFINDFLTDDANHAAYDIVAASNVAADRELDLEIHD 420
QY 421  ENQIPSRHSIQPQSGASSHGCGDGGIAGV 451
Db 421  ENQIPSRHSIQPQSGASSHGCGDGGIAGV 451

RESULT 5
US-09-816-277-16
; Sequence 16, Application US/09816277
; Publication No. US20010034045A1
; GENERAL INFORMATION:
; APPLICANT: Penttila, Merja E.
; APPLICANT: Ward, Michael
; APPLICANT: Wang, Huaming
; APPLICANT: Valkonen, Mari J.
; APPLICANT: Saloheimo, Markku
; TITLE OF INVENTION: Increased Production of Secreted
; FILE REFERENCE: GC590-2
; CURRENT APPLICATION NUMBER: US/09/816,277
; CURRENT FILING DATE: 2001-03-23
; PRIOR FILING DATE: 2000-03-24
; PRIOR APPLICATION NUMBER: US 09/534,692
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16
; LENGTH: 386
; TYPE: PRT
; ORGANISM: Aspergillus niger
US-10-663-450-6

Query Match      18.9%; Score 434.5; DB 4; Length 386;
Best Local Similarity 32.3%; Pred. No. 1.4e-20;
Matches 147; Conservative 50; Mismatches 139; Indels 119; Gaps 20;

QY 24  PGDNFTSLPADSTP-STLNPRDMTPDSVADIDS-----RLSVIPESQDAEDDES 72
Db 22  PGTTPGLAPPASPVNTRWPVFLMEEAFSPVDSLAGSPTPELPLLTVSPADTSLDSDS-- 79
QY 73  HSTSATAPSTSEKKPVKKRKSQGVLPPEKTNLPKRKAKTEDEKEORVERVLNRRAA 132
Db 80  -SVQAGETKAEKKPVKKRKSQGVLPPEKTNLPKRKAKTEDEKEORRTERVLNRRAA 138
QY 133  QSSRRERKRLVEALEKRNKELETLINVOKTNLILVEELNFRFRSSGVVTRSSSPLDSLQ 192
Db 139  QTSRRERKRLVEALEKRNKELETLINVOKTNLILVEELNFRFRSSGVVTRSSSPLDSLQ 192
QY 193  DSITLSQQLFGSRDQGTMSNPESQIMDSQIMRSAANPTVNPASLSPSLPISDKFQTEKE 252
Db 183  QVAQLSAEVRGSR-GNT---PKPG-----SPVSASPTLTPTLFKQ-----ER 220
QY 253  DE---EQADEEMEQTWHETKEAAAKKSKOSRVSTSTORPAVSIGG-----DAAVP 305
Db 221  DRIPLERIPFPTPTSTIDYSPTLRPSFLAB-----SSDVTQHPAVSVAGLEGEALS 272
QY 306  VFSDDAGANCLGLDPVHDDGPFSGHSGFLSAALDADRYLLESQLLASPNASTVDD- 362
Db 273  LF-DVGSN---PEPHAADD-----LAAPLSDDDFHRLFNVDSPVGSDDSVLEEDG 317
QY 363  ---DYLAGDSAACTNPLPSDYDFINDFLTDDANHAAYDIVAASNVAADRELDLEIHD 419
Db 318  FAFDVLDDGDLSAF--PPDSMWDFD-----PESVGFEGIE 350
QY 420  PENQIPSRHSIQ-----QPQSGASSHGCGDGGIAGV 450
Db 351  PPHGLPDETSRQTSSTSSVQPSLGASTRCDGQGGIAG 385

RESULT 6
US-10-663-450-16
; Sequence 16, Application US/10663450
; Publication No. US20040186070A1
; GENERAL INFORMATION:
; APPLICANT: Penttila, Merja E.
; APPLICANT: Ward, Michael
; APPLICANT: Wang, Huaming
; APPLICANT: Valkonen, Mari J.
; APPLICANT: Saloheimo, Markku
; TITLE OF INVENTION: Increased Production of Secreted
; FILE REFERENCE: GC590-2
; CURRENT APPLICATION NUMBER: US/10/663,450
; CURRENT FILING DATE: 2003-09-15
; PRIOR FILING DATE: 2000-03-24
; PRIOR APPLICATION NUMBER: US 09/534,692
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16
; LENGTH: 386
; TYPE: PRT
; ORGANISM: Aspergillus niger
US-10-663-450-16

Query Match      18.9%; Score 434.5; DB 4; Length 386;
Best Local Similarity 32.3%; Pred. No. 1.4e-20;
Matches 147; Conservative 50; Mismatches 139; Indels 119; Gaps 20;

QY 24  PGDNFTSLPADSTP-STLNPRDMTPDSVADIDS-----RLSVIPESQDAEDDES 72
Db 22  PGTTPGLAPPASPVNTRWPVFLMEEAFSPVDSLAGSPTPELPLLTVSPADTSLDSDS-- 79
```

```
QY 73 HSTSATAPSTSEKKPVKKRKSQGVLPKPTNLPPKRAKTEDEKEORRVERVLNRRAA 132
Db 80 -SVOAGETKAEKKPVKKRKSQGVLPKPTNLPPKRAKTEDEKEORRVERVLNRRAA 138
QY 133 QSRERKRLEVEALEKRNKELETLINVOXTNLIJLVEELNRRFRSSGVTVTRSSSPDLSIQ 192
Db 139 QTSRERKRLEMEKLENEK-----IQMEQONQFLQRLSQMEAEENRLN-----Q 182
QY 193 DSITLSQQLFGSRDGTMSNPEOSLMDQIMRSAANTVNPASISPSLPIPSDKEFQTKEE 252
Db 183 QVAQLSAEVRGR-GNT---PKPG-----SPVSGASPTLTPTLFKQ-----ER 220
QY 253 DE---EQADEMEQTHETKEAAAKKNSQSRVSTDSTQRPVAVSIGG-----DAAVP 305
Db 221 DEIPLERIPPTPSITDYSPTLRPSLAE-----SSDVTQHPAVSVAGLEGGSALS 272
QY 306 VFSDDAGANCLGLDPVHQDDGPFSGHSFGLSAAALDADRY--LLESQLLASPNASTVDP- 362
Db 273 LF--DVGSN---PEPHAADD-----LAAPLSDDDFHRLFNVDSPVGSDDSVLEDDG 317
QY 363 ---DYLAGDSAACTNPLPSDYPDINDFLTDDANHAAYDIVAASNAAADRELDLIHD 419
Db 318 FAFDVLDDGDLGAF--PFDSMVDFD-----PESVGFEGIE 350
QY 420 PENQIPSRHSIQ---QPSGASSHGCDGDIAGV 450
Db 351 PPHGLPDETSRQTSVQPSIGASTSRCDGGIAG 385

RESULT 7
US-09-816-277-4
; Sequence 4, Application US/09816277
; Publication No. US20010034045A1
; GENERAL INFORMATION:
; APPLICANT: Penttila, Merja E.
; APPLICANT: Ward, Michael
; APPLICANT: Wang, Huaming
; APPLICANT: Valkonen, Mari J.
; APPLICANT: Saloheimo, Markku
; TITLE OF INVENTION: Increased Production of Secreted
; TITLE OF INVENTION: Proteins by Recombinant Eukaryotic Cells
; FILE REFERENCE: GC590-2
; CURRENT APPLICATION NUMBER: US/09/816,277
; PRIOR FILING DATE: 2001-03-23
; PRIOR APPLICATION NUMBER: US 09/534,692
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 349
; TYPE: PRT
; ORGANISM: Aspergillus nidulans
US-09-816-277-4

Query Match 18.9%; Score 434; DB 3; Length 349;
Best Local Similarity 32.4%; Pred. No. 1.3e-20;
Matches 149; Conservative 50; Mismatches 109; Indels 152; Gaps 21;

QY 26 DNFTSL-----FADS--TPS-----TLNPRD--MMTPDSVADIDSRLSVIPESQDAEDD 70
Db 5 DRFSPVKMEDAFANSPPTPSLEVPVLTVPADTSLRTKNVQA-----47
QY 71 ESHSTSATAPSTSEKKPVKKRKSQGVLPKPTNLPPKRAKTEDEKEORRVERVLNR 130
Db 48 -----TKP--EEKKPAKKRKSQGVLPKPTNLPPKRAKTEDEKEORRVERVLNR 98
QY 131 AAQSRERKRLEVEALEKRNKELETLINVOXTNLIJLVEELNRRFRSSGVTVTRSSSPDLS 190
Db 99 AAQTSRERKRLEMEKLESEK-----IDMEQONQFLQRLAQMEAEENRLS-----143
QY 191 LQDSITLSQQLFGSRDGTMSNPEOSLMDQIMRSAANTVNPASISPSLPIPSDKEFQTK 250
Db 144 -QOVAQLSAEVRGRSRH---STPTSS-----SPASVSPTLTPTLFKQ-----180
QY 251 EDEDEQADE---DEMEQTHETKEAAAKKNSQSRVSTDSTQRPVAVSIGG-----DAA 303
Db 181 EGDEVPLDRIPPTPSVTDYSPTLPSSLAE-----SPDLTQHPAVSVGLEGDESA 232
QY 304 VPVFSDDAGANCLGLDPVHQDDGPFSS---IGHSFGLSAAALDADRYLLE-----SQ 350
Db 233 LTLF--DLGAS--IKHEPTHDLTAPLSDDDFRFLFNGDSSLESDDLLEGGFAFDVLDSGD 289
QY 351 LLASPNASTVDDYDLAGDSAACTNPLPSDYPDINDFLTDDANHAAYDIVAASNAAAD 410
```

```
Db 144 -QOVAQLSAEVRGRSRH---STPTSS-----SPASVSPTLTPTLFKQ-----180
QY 251 EDEDEQADE---DEMEQTHETKEAAAKKNSQSRVSTDSTQRPVAVSIGG-----DAA 303
Db 181 EGDEVPLDRIPPTPSVTDYSPTLPSSLAE-----SPDLTQHPAVSVGLEGDESA 232
QY 304 VPVFSDDAGANCLGLDPVHQDDGPFSS---IGHSFGLSAAALDADRYLLE-----SQ 350
Db 233 LTLF--DLGAS--IKHEPTHDLTAPLSDDDFRFLFNGDSSLESDDLLEGGFAFDVLDSGD 289
QY 351 LLASPNASTVDDYDLAGDSAACTNPLPSDYPDINDFLTDDANHAAYDIVAASNAAAD 410

RESULT 8
US-09-816-277-6
; Sequence 6, Application US/09816277
; Publication No. US20010034045A1
; GENERAL INFORMATION:
; APPLICANT: Penttila, Merja E.
; APPLICANT: Ward, Michael
; APPLICANT: Wang, Huaming
; APPLICANT: Valkonen, Mari J.
; APPLICANT: Saloheimo, Markku
; TITLE OF INVENTION: Increased Production of Secreted
; TITLE OF INVENTION: Proteins by Recombinant Eukaryotic Cells
; FILE REFERENCE: GC590-2
; CURRENT APPLICATION NUMBER: US/09/816,277
; PRIOR FILING DATE: 2001-03-23
; PRIOR APPLICATION NUMBER: US 09/534,692
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 349
; TYPE: PRT
; ORGANISM: Aspergillus nidulans
US-09-816-277-6

Query Match 18.9%; Score 434; DB 3; Length 349;
Best Local Similarity 32.4%; Pred. No. 1.3e-20;
Matches 149; Conservative 50; Mismatches 109; Indels 152; Gaps 21;

QY 26 DNFTSL-----FADS--TPS-----TLNPRD--MMTPDSVADIDSRLSVIPESQDAEDD 70
Db 5 DRFSPVKMEDAFANSPPTPSLEVPVLTVPADTSLRTKNVQA-----47
QY 71 ESHSTSATAPSTSEKKPVKKRKSQGVLPKPTNLPPKRAKTEDEKEORRVERVLNR 130
Db 48 -----TKP--EEKKPAKKRKSQGVLPKPTNLPPKRAKTEDEKEORRVERVLNR 98
QY 131 AAQSRERKRLEVEALEKRNKELETLINVOXTNLIJLVEELNRRFRSSGVTVTRSSSPDLS 190
Db 99 AAQTSRERKRLEMEKLESEK-----IDMEQONQFLQRLAQMEAEENRLS-----143
QY 191 LQDSITLSQQLFGSRDGTMSNPEOSLMDQIMRSAANTVNPASISPSLPIPSDKEFQTK 250
Db 144 -QOVAQLSAEVRGRSRH---STPTSS-----SPASVSPTLTPTLFKQ-----180
QY 251 EDEDEQADE---DEMEQTHETKEAAAKKNSQSRVSTDSTQRPVAVSIGG-----DAA 303
Db 181 EGDEVPLDRIPPTPSVTDYSPTLPSSLAE-----SPDLTQHPAVSVGLEGDESA 232
QY 304 VPVFSDDAGANCLGLDPVHQDDGPFSS---IGHSFGLSAAALDADRYLLE-----SQ 350
Db 233 LTLF--DLGAS--IKHEPTHDLTAPLSDDDFRFLFNGDSSLESDDLLEGGFAFDVLDSGD 289
QY 351 LLASPNASTVDDYDLAGDSAACTNPLPSDYPDINDFLTDDANHAAYDIVAASNAAAD 410
```

```
Db 290 LSAFFPDSVMDFD-----TEPVTLE-DLEQTNGLSASCKAASL----- 328
QY 411 RELDLEIHDPENQIPSRHSIQPQSGASSHGCDGGIAVG 450
Db 329 -----QPSHGASTSRCDDGGIAAG 347

RESULT 9
US-10-663-450-4
; Sequence 4, Application US/10663450
; Publication No. US20040186070A1
; GENERAL INFORMATION:
; APPLICANT: Penttila, Merja E.
; APPLICANT: Ward, Michael
; APPLICANT: Wang, Huaming
; APPLICANT: Valkonen, Mari J.
; APPLICANT: Saloheimo, Markku
; TITLE OF INVENTION: Increased Production of Secreted
; FILE REFERENCE: GCS90-2
; CURRENT APPLICATION NUMBER: US/10/663,450
; PRIOR FILING DATE: 2003-09-15
; PRIOR APPLICATION NUMBER: US 09/534,692
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 349
; TYPE: PRT
; ORGANISM: Aspergillus nidulans
US-10-663-450-4

Query Match 18.9%; Score 434; DB 4; Length 349;
Best Local Similarity 32.4%; Pred. No. 1.3e-20;
Matches 149; Conservative 50; Mismatches 109; Indels 152; Gaps 21;

QY 26 DNFTSL-----PADS-----TPS-----TLNPRD--MMTPDSVADIDSRLSVIPESQDAEDD 70
Db 5 DRFSPVKMEDAFANSPPTPSLEVPVLTSPADTSLRTKNVVAQ----- 47
QY 71 ESHSTSATAPSTSEKKPVKRSWGQVLPEPKTNLPKRKRAKTEDEKEORRVERVLNR 130
Db 48 -----TKP--EEKPAKRSWGQELVPKTNLPKRKRAKTEDEKEORRVERVLNR 98
QY 131 AAQSRERKRLEVEALEKRNKELETLINVKTNLILVEELNRRSSGVTSSPLDS 190
Db 99 AAQTSRERKRLEMEKLESEK-----IDMEQOQFLQLQAQMEANNRLS----- 143
QY 191 LQDSITLSQQLFGSRDQTMSPNPEOSLMDQIMRSAAANTVNPASLSPSLPISDKKEFQTK 250
Db 144 -QQVAQLSAEVRGSRH-----STPTSS-----SPASVSPTLTPTLFKQ----- 180
QY 251 BEDERQADE--DEMEQTHETKEAAAKENKSKQSRVSTSTORPAVSIIG-----DAA 303
Db 181 EGDEVPLDRIPPTPSVTDSYTLKPSLAE-----SPDLTQHPAVSVGGLGDESA 232
QY 304 VPVFSDDAGANCLGDPVHQDDGPFS---IGHSFGLSAAALDADRYLLE-----SQ 350
Db 233 LTLP--DLGAS-IXHEPTHDLTAPLSDDDFRFLFNGDSSLESDDLLEGGFAFDVLDSDG 289
QY 351 LLASPNASTVDDYLAGDSAACTNPLPSDYDFDINFLTDANHAAYDIVAASNYAAAD 410
Db 290 LSAFFPDSVMDFD-----TEPVTLE-DLEQTNGLSASCKAASL----- 328
QY 411 RELDLEIHDPENQIPSRHSIQPQSGASSHGCDGGIAVG 450
Db 329 -----QPSHGASTSRCDDGGIAAG 347
```

```
RESULT 10
US-10-663-450-6
; Sequence 6, Application US/10663450
```

```
; Publication No. US20040186070A1
; GENERAL INFORMATION:
; APPLICANT: Penttila, Merja E.
; APPLICANT: Ward, Michael
; APPLICANT: Wang, Huaming
; APPLICANT: Valkonen, Mari J.
; APPLICANT: Saloheimo, Markku
; TITLE OF INVENTION: Increased Production of Secreted
; FILE REFERENCE: GCS90-2
; CURRENT APPLICATION NUMBER: US/10/663,450
; CURRENT FILING DATE: 2003-09-15
; PRIOR APPLICATION NUMBER: US 09/534,692
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 349
; TYPE: PRT
; ORGANISM: Aspergillus nidulans
US-10-663-450-6

Query Match 18.9%; Score 434; DB 4; Length 349;
Best Local Similarity 32.4%; Pred. No. 1.3e-20;
Matches 149; Conservative 50; Mismatches 109; Indels 152; Gaps 21;

QY 26 DNFTSL-----PADS-----TPS-----TLNPRD--MMTPDSVADIDSRLSVIPESQDAEDD 70
Db 5 DRFSPVKMEDAFANSPPTPSLEVPVLTSPADTSLRTKNVVAQ----- 47
QY 71 ESHSTSATAPSTSEKKPVKRSWGQVLPEPKTNLPKRKRAKTEDEKEORRVERVLNR 130
Db 48 -----TKP--EEKPAKRSWGQELVPKTNLPKRKRAKTEDEKEORRVERVLNR 98
QY 131 AAQSRERKRLEVEALEKRNKELETLINVKTNLILVEELNRRSSGVTSSPLDS 190
Db 99 AAQTSRERKRLEMEKLESEK-----IDMEQOQFLQLQAQMEANNRLS----- 143
QY 191 LQDSITLSQQLFGSRDQTMSPNPEOSLMDQIMRSAAANTVNPASLSPSLPISDKKEFQTK 250
Db 144 -QQVAQLSAEVRGSRH-----STPTSS-----SPASVSPTLTPTLFKQ----- 180
QY 251 BEDERQADE--DEMEQTHETKEAAAKENKSKQSRVSTSTORPAVSIIG-----DAA 303
Db 181 EGDEVPLDRIPPTPSVTDSYTLKPSLAE-----SPDLTQHPAVSVGGLGDESA 232
QY 304 VPVFSDDAGANCLGDPVHQDDGPFS---IGHSFGLSAAALDADRYLLE-----SQ 350
Db 233 LTLP--DLGAS-IXHEPTHDLTAPLSDDDFRFLFNGDSSLESDDLLEGGFAFDVLDSDG 289
QY 351 LLASPNASTVDDYLAGDSAACTNPLPSDYDFDINFLTDANHAAYDIVAASNYAAAD 410
Db 290 LSAFFPDSVMDFD-----TEPVTLE-DLEQTNGLSASCKAASL----- 328
QY 411 RELDLEIHDPENQIPSRHSIQPQSGASSHGCDGGIAVG 450
Db 329 -----QPSHGASTSRCDDGGIAAG 347

RESULT 11
US-09-533-029-104
; Sequence 104, Application US/09533029
; Publication No. US20030046723A1
; GENERAL INFORMATION:
; APPLICANT: Heard, Jacqueline
; APPLICANT: Broun, Pierre
; APPLICANT: Riechmann, Jose-Luis
; APPLICANT: Keddies, James
; APPLICANT: Pineda, Omaira
; APPLICANT: Adam, Luc
; APPLICANT: Samaha, Raymond
; APPLICANT: Zhang, James
; APPLICANT: Yu, Guo-Liang
```

```

; APPLICANT: Ratcliffe, Oliver
; APPLICANT: Pilgrim, Marsha
; APPLICANT: Jiang, Cai-Zhong
; APPLICANT: Reuber, Lynne
; TITLE OF INVENTION: DISEASE-INDUCED POLYNUCLEOTIDES
; FILE REFERENCE: MBI-010
; CURRENT APPLICATION NUMBER: US/09/533,029
; CURRENT FILING DATE: 2000-03-22
; EARLIER APPLICATION NUMBER: 60/125,814
; EARLIER FILING DATE: 1999-03-23
; NUMBER OF SEQ ID NOS: 121
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 104
; LENGTH: 409
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; OTHER INFORMATION: G1034
; US-09-533-029-104

Query Match      18.7%; Score 429.5; DB 3; Length 409;
Best Local Similarity 38.2%; Pred. No. 3.2e-20;
Matches 116; Conservative 39; Mismatches 70; Indels 79; Gaps 10

Qy      11 KFEASPA-----ESFLSAPGDNFTSLFADSTPSTLNPRDMTPDSDVADISRLSVIPESQD 66
Db      8 KYENSPAETMVESFVSTPSSPHNPLFD---NNLNPVDGFGSP-----QS 48

Qy      67 ABDESHSTATATSESTSEKKVPKKKSGOVLPEPKTNLPKRKRAKTEDEKEORRVERVL 126
Db      49 FDRDYNFNGLSLGNLPEKPIKRRKSGQQLPPEKTNLPKRKRAKTOQEKEQRRVERVL 108

Qy      127 RNRRAAGSSRRERKRLEVAL-----EKRKKELETLINVOKTNLIILVEELNR---FRR 176
Db      109 RNRRAAGSSRRERKRQVEALEVEKRAIERKKNWDLMEWLADMEAKYLLQOELKRASGYNK 168

Qy      177 SSGVVTRSSSPSLDLSQDS-----ITLSQQLFGSRDQGTMSNPQSLMDQIMRSAANPTVNP 232
Db      169 TNFLSYSDSPDISEDQLSPLTFSKQLFNAQD--ELCRP---ISPQSIGPLTSTRTVDP 223

Qy      233 ASLSPSLPPISDKEFQTKEDDEEQADEDEMEQTHETKEAAAAAEEKSKQSRFVSTDSTQ 292
Db      224 STLSP-----KSLSSPDSSNSN-----SSDMTQ 246

Qy      293 RPAV 296
Db      247 HPAV 250

RESULT 12
US-10-295-403-158
; Sequence 158, Application US/10295403
; Publication No. US20030101481A1
; GENERAL INFORMATION:
; APPLICANT: Heard, Jacqueline
; APPLICANT: Riechmann, Jose Luis
; APPLICANT: Adam, Luc
; APPLICANT: Broun, Pierre
; APPLICANT: Pineda, Omaira
; APPLICANT: Reuber, Lynne
; APPLICANT: Jiang, Cai-Zhong
; APPLICANT: Keddie, James
; APPLICANT: Zhang, James
; APPLICANT: Benito, Maria-Ines
; APPLICANT: Yu, Guo-Liang
; APPLICANT: Fromm, Mike
; TITLE OF INVENTION: PLANT GENE SEQUENCES I
; FILE REFERENCE: MBI-0003
; CURRENT APPLICATION NUMBER: US/10/295,403
; CURRENT FILING DATE: 2002-11-15
; PRIOR APPLICATION NUMBER: US/09/394,519
; PRIOR FILING DATE: 1999-09-13
; PRIOR APPLICATION NUMBER: 60/101,349

```

```

; CURRENT FILING DATE: 2003-04-10
; PRIOR APPLICATION NUMBER: 09/394,519
; PRIOR FILING DATE: 1999-09-13
; PRIOR APPLICATION NUMBER: 09/489,376
; PRIOR FILING DATE: 2000-01-21
; PRIOR APPLICATION NUMBER: 09/506,720
; PRIOR FILING DATE: 2000-02-17
; PRIOR APPLICATION NUMBER: 09/533,030
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: 09/533,392
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: 09/533,029
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: 09/532,591
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: 09/533,648
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: 09/713,994
; PRIOR FILING DATE: 2000-11-16
; PRIOR APPLICATION NUMBER: 09/819,142
; PRIOR FILING DATE: 2001-03-27
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 2011
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 684
; LENGTH: 409
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; OTHER INFORMATION: G1034
US-10-412-699B-684
```

```

Query Match      18.7%; Score 429.5; DB 4; Length 409;
Best Local Similarity 38.2%; Pred. No. 3.2e-20;
Matches 116; Conservative 39; Mismatches 70; Indels 79; Gaps 10;

Qy 11 KFEASPA---ESFLSAPGDNFTSLFADSTSTLNPRDMWTPDSVADISRLSVIPESQD 66
Db 8 KYENSPASTWSEFVSTSSFNPLFD---NNLNPVDGFSF-----QS 48

Qy 67 AEDDESHTSATAPSTSEKKPVKKKSGQVLPPEKTNLPKRKRAKTEDEKEQRRVERVL 126
Db 49 FDRDYNFNGSLSGNLPEKKIPEKKKSGQVLPPEKTNLPKRKRAKTEDEKEQRRVERVL 108

Qy 127 RNRAAQSRRKRLEVEAL-----EKRKKELETLINVKTNLILVEELNR----FRR 176
Db 109 RNRAAQSRRKRQVEALEKRAIERKNMDLEMLADMEAKYLLQQLKRAAGYNYK 168

Qy 177 SSGVVTRSSPLDSLQDS-----ITLSQQLFGSRDQGTMSNPEQSLMDQIMRSAANTVNP 232
Db 169 TNFLSYSDSTPOISEDSQLSPLTFSKQLFNAQD--ELCRP---ISPSIGLTSRTVDP 223

Qy 233 ASLSPSLPPISDKEFOTKEEBEQADEMEQTHETKEAAAKKSKQSRVSTDSQ 292
Db 224 STLSP-----KSLSSPDSSNSN-----SSDWTQ 246

Qy 293 RPAV 296
Db 247 HPAV 250
```

```

RESULT 14
US-09-816-277-19
; Sequence 19, Application US/09816277
; Publication No. US20010034045A1
; GENERAL INFORMATION:
; APPLICANT: Penttila, Merja E.
; APPLICANT: Ward, Michael
; APPLICANT: Wang, Huaming
; APPLICANT: Valkonen, Mari J.
; APPLICANT: Saloheimo, Markku
; TITLE OF INVENTION: Increased Production of Secreted
; TITLE OF INVENTION: Proteins by Recombinant Eukaryotic Cells
```

```

; FILE REFERENCE: GC590-2
; CURRENT APPLICATION NUMBER: US/09/816,277
; CURRENT FILING DATE: 2001-03-23
; PRIOR APPLICATION NUMBER: US 09/534,692
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 19
; LENGTH: 342
; TYPE: PRT
; ORGANISM: Aspergillus niger
US-09-816-277-19

Query Match      18.6%; Score 426.5; DB 3; Length 342;
Best Local Similarity 31.7%; Pred. No. 4.1e-20;
Matches 145; Conservative 44; Mismatches 125; Indels 143; Gaps 20;

Qy 15 SPAESFLSAPGDNFTSLFADSTPS---TLAPRDMWTPDSVADISRLSVIPESQDAEDD 70
Db 7 SPVDSLAGSP-----TPELPLLTVPADTSLDSS----- 35

Qy 71 ESHSTSATAPSTSEKKPVKKKSGQVLPPEKTNLPKRKRAKTEDEKEQRRVERVLNR 130
Db 36 ---SVQAGETKAEKKPVKKKSGQELVPKTNLPKRKRAKTEDEKEQRRVERVLNR 92

Qy 131 AAQSRRKRLEVEALEKRNKELETLINVKTNLILVEELNRPRSSGVVTRSSPLDS 190
Db 93 AAQTSRRKRLEMEKLENEK-----IQMEQQNQFLQLRSLQMEANRLN----- 137

Qy 191 LQDSITLSQQLFGSRDQGTMSNPEQSLMDQIMRSAANTVNPASLSPSLPISDKKEFQTK 250
Db 138 -QQVAQLSAEVGRSGR-GNT---PKPG-----SPVSASPLTPTLFLFKQ---- 174

Qy 251 BEDE---EQADEDEMEQTHETKEAAAKKSKQSRVSTDSQRPVSIIGG---DAA 303
Db 175 ERDIPLERIFPPPTSIDYSPTLRPSTLAE-----SSDVTQHPAVSVAGLEGE 226

Qy 304 VPVFSDDAGANCLGDPVHQDGPFSIGHSGLSAALDADRY--LLESQLLASPNASTVD 361
Db 227 LSLF--DVGSN---PEPHAADD-----LAAPLSDDDFHRLFNVDSPVGSDDSVLE 271

Qy 362 D---DVLGDSAACTNPLPSDYDFDINDFLTDDANHAAYDIVAASNYAAADRELDLEI 417
Db 272 DGFAFDVLDGDLGSAF--PFDSMVDFD-----PESVGFEG 304

Qy 418 HDPENQIPSRHSIQ-----QPOSQASSHGCDGDIAGV 450
Db 305 IEPHGLPDETSRGTSSVQPSLGSASTSRCDGQGIAG 341
```

```

RESULT 15
US-10-663-450-19
; Sequence 19, Application US/10663450
; Publication No. US20040186070A1
; GENERAL INFORMATION:
; APPLICANT: Penttila, Merja E.
; APPLICANT: Ward, Michael
; APPLICANT: Wang, Huaming
; APPLICANT: Valkonen, Mari J.
; APPLICANT: Saloheimo, Markku
; TITLE OF INVENTION: Increased Production of Secreted
; TITLE OF INVENTION: Proteins by Recombinant Eukaryotic Cells
; FILE REFERENCE: GC590-2
; CURRENT APPLICATION NUMBER: US/10/663,450
; CURRENT FILING DATE: 2003-09-15
; PRIOR APPLICATION NUMBER: US 09/534,692
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 19
; LENGTH: 342
; TYPE: PRT
; ORGANISM: Aspergillus niger
```

Sat Nov 26 13:50:38 2005

US-10-663-450-19

```
Query Match      18.6%; Score 426.5; DB 4; Length 342;
Best Local Similarity 31.7%; Pred. No. 4.1e-20;
Matches 145; Conservative 44; Mismatches 125; Indels 143; Gaps 20;

Qy      15 SPAESFLSAPGDNFTSLFADSTPS---TLNPRDMTTPDSVADIDSRLSVIPESQDAEDD 70
Db      7 SPVDSLAGSP-----TPELPLTVSPADTSLDDS-----35

Qy      71 ESHSTATAPSTSKKPKKKGKQVLPKPKTNLPKRAKTEDEKEORVERVLNRR 130
Db      36 ---SVQAGETKAEEKKPKKKGKQVLPKPKTNLPKRAKTEDEKEORRIERVLRRA 92

Qy      131 AAQSSRRERKLEVALEKRNKELETLINVKTNLILVEELNRRFRSSGVVTRSSPLDS 190
Db      93 AAQTSRRERKLEMEKLENEK-----IQMEQQNQFLLQRLSQMEANNRLN-----137

Qy      191 LQDSITLSQLFGSRDQGTMSNPQSLMDQIMRSAANPTVNPASLSPSLPISDKFQTK 250
Db      138 -QQVAQLSAEVRGSR-GNT---PKPG-----SPVSASPTLTPTLFKQ----174

Qy      251 EEDE---EQADEDEMEQTWETKEAAAKKSKQSRVSTDTSTORPAVSIGG-----DAA 303
Db      175 ERDEIPLERIPTPTSIPTDYSPTLRPSTLAE-----SSDVTQHPAVSVAGLEGGSA 226

Qy      304 VPVFSDDAGANCLGLDPVHQDDGPFSGHSGLSAALDADRY--LLESQLLASPNASTVD 361
Db      227 LSLF--DVGSN---PEPHAADD-----LAAPLSDDDFHRLFNVDSPVGSDDSVLE 271

Qy      362 D----DYLAGDSACFTNPLPFSYDFDINDFLTDANHAAYDIVAASNAAADRELDLEI 417
Db      272 DGFAFDVLDGGLSAF--PFDSMWDFD-----PESVGFEG 304

Qy      418 HDPENQIPSRHSIQ-----QPSGASHGCGDDGGIAGV 450
Db      305 IEPHGLPDETSRQTSSVQPSLGASTSRCGQGIAG 341
```

Search completed: November 23, 2005, 03:32:19
Job time : 148.003 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 23, 2005, 03:18:52 ; Search time 0.861751 Seconds
(without alignments)
239.171 Million cell updates/sec

Title: US-10-663-450-60

Perfect score: 342

Sequence: 1 KSTLPKRAKTKKEKQRR.....CSLLENLNSVNLKLADEH 68

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 17545 seqs, 3030971 residues

Total number of hits satisfying chosen parameters: 17545

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA New.*
1: /cgn2_6/prodata/2/pubpa/US10_NEW_PUB_PEP.*
2: /cgn2_6/prodata/2/pubpa/US06_NEW_PUB_PEP.*
3: /cgn2_6/prodata/2/pubpa/US07_NEW_PUB_PEP.*
4: /cgn2_6/prodata/2/pubpa/US08_NEW_PUB_PEP.*
5: /cgn2_6/prodata/2/pubpa/US09_NEW_PUB_PEP.*
6: /cgn2_6/prodata/2/pubpa/PCT_NEW_PUB_PEP.*
7: /cgn2_6/prodata/2/pubpa/US11_NEW_PUB_PEP.*
8: /cgn2_6/prodata/2/pubpa/US60_NEW_PUB_PEP.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	87.5	25.6	135	1	US-10-821-234-1037
2	72	21.1	454	7	US-11-169-013-1
3	69	20.2	376	1	US-10-821-234-1037
4	63	18.4	579	1	US-10-821-234-874
5	62.5	18.3	431	7	US-11-169-013-2
6	61.5	18.0	86	1	US-10-667-295-128
7	61.5	18.0	97	1	US-10-667-295-127
8	61.5	18.0	431	1	US-10-821-234-1285
9	60	17.3	383	1	US-10-793-626-3332
10	59	17.3	453	1	US-10-131-826A-270
11	59	17.3	457	1	US-10-821-234-1185
12	58.5	17.1	1388	1	US-10-821-234-1143
13	56	16.4	625	1	US-10-310-386-6
14	56	16.4	776	1	US-10-821-234-1171
15	55.5	16.2	82	1	US-10-667-295-69
16	55.5	16.2	90	1	US-10-667-295-67
17	55.5	16.2	185	1	US-10-821-234-1130
18	55.5	16.2	586	1	US-10-131-826A-46
19	55	16.1	112	7	US/11/020
20	55	16.1	284	1	US-10-821-234-1632
21	54.5	15.9	639	1	US-10-821-234-907
22	54.5	15.9	2432	1	US-10-821-234-899
23	54	15.8	401	1	US-10-821-234-881
24	53.5	15.6	88	1	US-10-667-295-190
25	53.5	15.6	1614	1	US-10-821-234-903

26 53 15.5 171 1 US-10-821-234-994 Sequence 994, Appli
27 52.5 15.4 128 7 US-11-105-864-4 Sequence 4, Appli
28 52.5 15.4 578 1 US-10-821-234-1039 Sequence 1039, Ap
29 52 15.2 267 1 US-10-793-626-882 Sequence 882, Appli
30 52 15.2 400 7 US-11-018-018-7 Sequence 7, Appli
31 52 15.2 400 7 US-11-047-757-7 Sequence 2, Appli
32 52 15.2 747 7 US-11-182-592-2 Sequence 7, Appli
33 51.5 15.1 163 1 US-10-689-742-42 Sequence 42, Appli
34 51.5 15.1 475 1 US-10-793-626-2076 Sequence 2076, Ap
35 51.5 15.1 525 1 US-10-793-626-1292 Sequence 1292, Ap
36 51.5 15.1 559 1 US-10-821-234-1513 Sequence 1513, Ap
37 51.5 15.1 719 1 US-10-793-626-1548 Sequence 1548, Ap
38 51 14.9 322 1 US-10-793-626-1990 Sequence 1990, Ap
39 51 14.9 877 7 US-11-074-176-322 Sequence 322, App
40 51 14.9 883 7 US-11-074-176-88 Sequence 88, Appli
41 51 14.9 1616 1 US-10-821-234-1497 Sequence 1497, Ap
42 50.5 14.8 122 1 US-10-667-295-177 Sequence 177, App
43 50.5 14.8 1059 7 US/11/062 Sequence 5, Appli
44 50.5 14.8 1084 7 US/11/062 Sequence 8, Appli
45 50 14.6 410 1 US-10-821-234-1180 Sequence 1180, Ap

ALIGNMENTS

RESULT 1

US-10-821-234-1037
; Sequence 1037, Application US/10821234
; Publication No. US20050255114A1
; GENERAL INFORMATION:
; APPLICANT: Labat, Ivan
; APPLICANT: Stache-Crain, Birgit
; APPLICANT: Andarmani, Susan
; APPLICANT: Tang, Y. Tom
; TITLE OF INVENTION: Methods for Diagnosis and Treatment of Preeclampsia
; FILE REFERENCE: 821A
; CURRENT APPLICATION NUMBER: US/10/821,234
; PRIOR FILING DATE: 2004-04-07
; PRIOR APPLICATION NUMBER: US 60/462,047
; PRIOR FILING DATE: 2003-04-07
; NUMBER OF SEQ ID NOS: 1704
; SOFTWARE: pt_seq_genes Version 1.0
; SEQ ID NO 1037
; LENGTH: 135
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-821-234-1037

Query Match 25.6%; Score 87.5; DB 1; Length 135;
Best Local Similarity 34.5%; Pred. No. 0.002;
Matches 20; Conservative 16; Mismatches 21; Indels 1; Gaps 1;

QY 7 RKRATKKEKQRRIRLRNRRAHQSRKKRLHLQYLERKCSLLENLNSVNLK 64
DB 67 RSFQLAEEATRKRELRLMKNRRAECCRKKYKCLNRVAVLEN-QNKTLEEL 123

RESULT 2

US-11-169-013-1
; Sequence 1, Application US/11169013
; Publication No. US20050244971A1
; GENERAL INFORMATION:
; APPLICANT: Korea Kumho Petrochemical Co., Ltd.
; TITLE OF INVENTION: TRANSGENIC PLANTS WITH ENHANCED STRESS TOLERANCE
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/11/169,013
; CURRENT FILING DATE: 2005-06-29
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: KopatentIn 1.71
; SEQ ID NO 1
; LENGTH: 454
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana

```

US-11-169-013-1
Query Match      21.1%; Score 72; DB 7; Length 454;
Best Local Similarity 32.7%; Pred. No. 0.33;
Matches 16; Conservative 14; Mismatches 19; Indels 0; Gaps 0;

QY   7 RKRAKTKKEQRRIERILNRRAAHQSREKKRLHLQYLKERKCSLLENL 55
Db   362 RKTGAVLEKVIERRQRMKNRESAARSARKQAAYTMELEAEIAQLKEL 410

RESULT 3
US-10-971-994-2
; Sequence 2, Application US/10971994
; Publication No. US20050250182A1
; GENERAL INFORMATION:
; APPLICANT: University of Michigan et al.
; APPLICANT: Kaufman, Randal
; APPLICANT: Kyungo, Lee
; APPLICANT: Mori, Kazutos
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, AND METHODS FOR
; TITLE OF INVENTION: MODULATING THE UNFOLDED PROTEIN RESPONSE
; FILE REFERENCE: UMW-2246
; CURRENT APPLICATION NUMBER: US/10/971,994
; CURRENT FILING DATE: 2004-10-21
; PRIOR APPLICATION NUMBER: PCT/2003/012640
; PRIOR FILING DATE: 2003-04-22
; PRIOR APPLICATION NUMBER: 60/375,098
; PRIOR FILING DATE: 2002-04-22
; PRIOR APPLICATION NUMBER: 60/374,880
; PRIOR FILING DATE: 2002-04-23
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 376
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-971-994-2

Query Match      20.2%; Score 69; DB 1; Length 376;
Best Local Similarity 33.3%; Pred. No. 0.55;
Matches 20; Conservative 12; Mismatches 18; Indels 10; Gaps 2;

QY   7 RKRAK-TKEEQRRIERILNRRAAHQSREKKRLHLQYLKERKC-----SLLENLL 56
Db   59 RKQRLTHLSPEKALRRKLKRVAAQTARDKKARMSLEQQVVVDLEENQKLLLENQL 118

RESULT 4
US-10-821-234-874
; Sequence 874, Application US/10821234
; Publication No. US20050255114A1
; GENERAL INFORMATION:
; APPLICANT: Labat, Ivan
; APPLICANT: Stache-Crain, Birgit
; APPLICANT: Andarmani, Susan
; APPLICANT: Tang, Y. Tom
; TITLE OF INVENTION: Methods for Diagnosis and Treatment of Preeclampsia
; FILE REFERENCE: 821A
; CURRENT APPLICATION NUMBER: US/10/821,234
; CURRENT FILING DATE: 2004-04-07
; PRIOR APPLICATION NUMBER: US 60/462,047
; PRIOR FILING DATE: 2003-04-07
; NUMBER OF SEQ ID NOS: 1704
; SOFTWARE: pt_seq_genes Version 1.0
; SEQ ID NO 874
; LENGTH: 579
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-821-234-874

Query Match      18.4%; Score 63; DB 1; Length 579;
Best Local Similarity 45.5%; Pred. No. 3.8;
Matches 15; Conservative 4; Mismatches 14; Indels 0; Gaps 0;

QY   7 RKRAKTKKEQRRIERILNRRAAHQSREKKRLHLQYLKERKCSLLENL 55
Db   362 RKTGAVLEKVIERRQRMKNRESAARSARKQAAYTMELEAEIAQLKEL 410

RESULT 5
US-11-169-013-2
; Sequence 2, Application US/11169013
; Publication No. US20050244971A1
; GENERAL INFORMATION:
; APPLICANT: Korea Kumho Petrochemical Co., Ltd.
; TITLE OF INVENTION: TRANSGENIC PLANTS WITH ENHANCED STRESS TOLERANCE
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/11/169,013
; CURRENT FILING DATE: 2005-06-29
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: KopatentIn 1.71
; SEQ ID NO 2
; LENGTH: 431
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-11-169-013-2

Query Match      18.3%; Score 62.5; DB 7; Length 431;
Best Local Similarity 34.0%; Pred. No. 3.1;
Matches 16; Conservative 13; Mismatches 17; Indels 1; Gaps 1;

QY   8 KRAKTKKEKE-QRRIERILNRRAAHQSREKKRLHLQYLKERKCSLLE 53
Db   341 RSSNTGLEKVIERRQRMKNRESAARSARKQAAYTLEAEIEKLK 387

RESULT 6
US-10-667-295-128
; Sequence 128, Application US/10667295
; Publication No. US20050257293A1
; GENERAL INFORMATION:
; APPLICANT: Mascia, Peter
; TITLE OF INVENTION: BIOLOGICAL CONTAINMENT SYSTEM
; FILE REFERENCE: 11696-047001
; CURRENT APPLICATION NUMBER: US/10/667,295
; CURRENT FILING DATE: 2003-09-17
; PRIOR APPLICATION NUMBER: US 60/411,823
; PRIOR FILING DATE: 2002-09-17
; NUMBER OF SEQ ID NOS: 263
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 128
; LENGTH: 86
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(86)
; OTHER INFORMATION: Ceres Seq. ID no. 12438738
US-10-667-295-128

Query Match      18.0%; Score 61.5; DB 1; Length 86;
Best Local Similarity 32.3%; Pred. No. 0.6;
Matches 21; Conservative 14; Mismatches 25; Indels 5; Gaps 3;

QY   8 KRAKTKKEQRRIERILNR--RRAAHQSREKKRLHLQYLKERKS--LLENLN--SVNLE 62
Db   14 RAARSKSEERCKIESIISQFRNAEDVDRKNFLYEYLLLRGNKOLDQLKPPTGLS 73

QY   63 KLADH 67
Db   74 SLQLH 78

RESULT 7
US-10-667-295-127

```



```
; Sequence 127, Application US/10667295
; Publication No. US20050257293A1
; GENERAL INFORMATION:
; APPLICANT: Mascia, Peter
; TITLE OF INVENTION: BIOLOGICAL CONTAINMENT SYSTEM
; FILE REFERENCE: 11696-047001
; CURRENT APPLICATION NUMBER: US/10/667,295
; PRIOR FILING DATE: 2003-09-17
; PRIOR APPLICATION NUMBER: US 60/411,823
; PRIOR FILING DATE: 2002-09-17
; NUMBER OF SEQ ID NOS: 263
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 127
; LENGTH: 97
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(97)
; OTHER INFORMATION: Ceres Seq. ID no. 12438737
US-10-667-295-127

Query Match      18.0%; Score 61.5; DB 1; Length 97;
Best Local Similarity 32.3%; Pred. No. 0.69;
Matches 21; Conservative 14; Mismatches 25; Indels 5; Gaps 3;

QY      8 KRAKTKKEEKQRIERILRN--RRAHQREKRLHLQYLERKCS-LLENLLN--SVNLE 62
DB      25 RAARSKSEBERCKIESIISQEFRNAEDVDKRNFLYIEVLLRRGNQLDQKNGPTTGLS 84
QY      63 KLAHD 67
DB      85 SLQLH 89

RESULT 8
US-10-821-234-1285
; Sequence 1285, Application US/10821234
; Publication No. US20050255114A1
; GENERAL INFORMATION:
; APPLICANT: Labat, Ivan
; APPLICANT: Stache-Crain, Birgit
; APPLICANT: Andarmani, Susan
; APPLICANT: Tang, Y. Tom
; TITLE OF INVENTION: Methods for diagnosis and Treatment of Preeclampsia
; FILE REFERENCE: 821A
; CURRENT APPLICATION NUMBER: US/10/821,234
; PRIOR FILING DATE: 2004-04-07
; PRIOR APPLICATION NUMBER: US 60/462,047
; PRIOR FILING DATE: 2003-04-07
; NUMBER OF SEQ ID NOS: 1704
; SOFTWARE: pt_seq_genes Version 1.0
; SEQ ID NO 1285
; LENGTH: 431
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-821-234-1285

Query Match      18.0%; Score 61.5; DB 1; Length 431;
Best Local Similarity 32.2%; Pred. No. 3.9;
Matches 19; Conservative 14; Mismatches 21; Indels 5; Gaps 2;

QY      8 KRAKTK-----EKEQRRIERILNRRAHQREKRLHLQYLERKCSLENNLSVNL 62
DB      176 RRGKVEQLSPBEKEKRIRRR-ERNKMAAKCRNRRLTDTLQAEYDQLEDEKSAEQTE 233

RESULT 9
US-10-793-626-3332
; Sequence 3332, Application US/10793626
; Publication No. US20050255478A1
; GENERAL INFORMATION:
; APPLICANT: KIMMERLY, WILLIAM JOHN
```

```
; TITLE OF INVENTION: STAPHYLOCOCCUS EPIDERMIDIS NUCLEIC ACIDS AND PROTEINS
; FILE REFERENCE: PU34800S
; CURRENT APPLICATION NUMBER: US/10/793,626
; CURRENT FILING DATE: 2004-03-04
; PRIOR APPLICATION NUMBER: 60/164,258
; PRIOR FILING DATE: 1999-11-09
; NUMBER OF SEQ ID NOS: 4472
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 3332
; LENGTH: 383
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic
; OTHER INFORMATION: amino acid sequence
US-10-793-626-3332

Query Match      17.5%; Score 60; DB 1; Length 383;
Best Local Similarity 32.8%; Pred. No. 4.9;
Matches 19; Conservative 11; Mismatches 24; Indels 4; Gaps 2;

QY      7 RRAKTKKEEKQRIERI--LRNRAAHOQREKRLHLQY--LERKCSLENNLSVNV 60
DB      28 RAGAKARTTKQQAIRNRFNQLSDVKTQHTQDKGELNLAYSRLGKQVYELKNLSKSN 85

RESULT 10
US-10-131-826A-270
; Sequence 270, Application US/10131826A
; Publication No. US20050245730A1
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeForge, Laura
; APPLICANT: Desnoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330RIC128
; CURRENT APPLICATION NUMBER: US/10/131,826A
; CURRENT FILING DATE: 2002-04-24
; PRIOR APPLICATION NUMBER: 60/049911
; PRIOR FILING DATE: 1997-06-18
; PRIOR APPLICATION NUMBER: 60/056974
; PRIOR FILING DATE: 1997-08-26
; PRIOR APPLICATION NUMBER: 60/059113
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059115
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059117
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059122
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059184
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059263
; PRIOR FILING DATE: 1997-09-18
; PRIOR APPLICATION NUMBER: 60/059352
; PRIOR FILING DATE: 1997-09-19
; PRIOR APPLICATION NUMBER: 60/059588
; PRIOR FILING DATE: 1997-09-19
```


Query Match 16.4%; Score 56; DB 1; Length 776;
Best Local Similarity 27.3%; Pred. No. 29;
Matches 18; Conservative 11; Mismatches 21; Indels 16; Gaps 2;
Qy 14 BEKEQRRIERI-----LRNRRAAHQGREKKRLHLQYLERKCSLLENLLNSVNL 62
Db 479 QAEQDKIEVRDAMENEMRTQLRQAAATDHLRDVLRVQEQELKSEFEQNL-----SE 533
Qy 63 KLDHE 68
Db 534 KLSEQE 539

RESULT 15
US-10-667-295-69
; Sequence 69, Application US/10667295
; Publication No. US2005257293A1
; GENERAL INFORMATION:
; APPLICANT: Mascia, Peter
; TITLE OF INVENTION: BIOLOGICAL CONTAINMENT SYSTEM
; FILE REFERENCE: 11696-047001
; CURRENT APPLICATION NUMBER: US/10/667,295
; CURRENT FILING DATE: 2003-09-17
; PRIOR APPLICATION NUMBER: US 60/411,823
; PRIOR FILING DATE: 2002-09-17
; NUMBER OF SEQ ID NOS: 263
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 69
; LENGTH: 82
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(82)
; OTHER INFORMATION: Ceres Seq. ID no. 12561624
US-10-667-295-69

Query Match 16.2%; Score 55.5; DB 1; Length 82;
Best Local Similarity 30.8%; Pred. No. 2.4;
Matches 12; Conservative 13; Mismatches 13; Indels 1; Gaps 1;
Qy 8 KRAKTKBEKEQRRIERIILNRRAAHQGREKKRLHLQYLE 46
Db 14 RAARSPIEDRRRIEMIV-STEFHNSKEVDKKNFYIE 51

Search completed: November 23, 2005, 03:32:38
Job time : 0.861751 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 23, 2005, 03:12:56 ; Search time 22.0138 Seconds
(without alignments)
1290.661 Million cell updates/sec

Title: US-10-663-450-60

Perfect score: 342

Sequence: 1 KSTLPPRKAKTKEKEQRR.....CSLLENLNSVNEKLADHE 68

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA.Main:

- 1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pap:*
- 2: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pap:*
- 3: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pap:*
- 4: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pap:*
- 5: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pap:*
- 6: /cgn2_6/ptodata/1/pubpaa/US11_PUBCOMB.pap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	342	100.0	68	3	US-09-816-277-60
2	342	100.0	68	4	US-10-663-450-60
3	342	100.0	200	4	US-10-369-493-21867
4	196	57.3	451	3	US-09-816-277-2
5	196	57.3	451	3	US-09-816-277-5
6	196	57.3	451	4	US-10-663-450-5
7	196	57.3	451	4	US-10-663-450-5
8	178.5	52.2	409	3	US-09-816-277-104
9	178.5	52.2	409	4	US-10-395-403-158
10	178.5	52.2	409	4	US-10-412-699B-684
11	176.5	51.6	342	3	US-09-816-277-19
12	176.5	51.6	342	4	US-10-663-450-19
13	176.5	51.6	349	3	US-09-816-277-4
14	176.5	51.6	349	3	US-09-816-277-6
15	176.5	51.6	349	4	US-10-663-450-4
16	176.5	51.6	349	4	US-10-663-450-6
17	176.5	51.6	386	3	US-09-816-277-16
18	176.5	51.6	386	4	US-10-663-450-16
19	171.5	50.1	174	4	US-10-767-701-61238
20	124.5	36.4	168	3	US-09-934-455-102
21	124.5	36.4	168	4	US-10-225-066A-206
22	124.5	36.4	168	4	US-10-374-780A-2356
23	124.5	36.4	168	5	US-10-732-923-13692
24	124.5	36.4	168	5	US-10-225-066A-206
25	122	35.7	203	4	US-10-437-963-151695
26	122	35.7	203	5	US-10-732-923-13694
27	122	35.7	203	5	US-10-732-923-13695

US-09-816-277-60
; Sequence 60, Application US/09816277
; Publication No. US20010034045A1
; GENERAL INFORMATION:
; APPLICANT: Penttila, Merja E.
; APPLICANT: Ward, Michael
; APPLICANT: Wang, Huaming
; APPLICANT: Valkonen, Mari J.
; APPLICANT: Saloheimo, Markku
; TITLE OF INVENTION: Increased Production of Secreted
; TITLE OF INVENTION: Proteins by Recombinant Eukaryotic Cells
; FILE REFERENCE: GC590-2
; CURRENT APPLICATION NUMBER: US/09/816, 277
; CURRENT FILING DATE: 2001-03-23
; PRIOR APPLICATION NUMBER: US 09/534, 692
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 60
; LENGTH: 68
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-09-816-277-60

US-10-732-923-13697
Sequence 2, Appli
Sequence 2, Appli
Sequence 6, Appli
Sequence 10, Appli
Sequence 2, Appli
Sequence 6, Appli
Sequence 10, Appli
Sequence 197786,
Sequence 585, App
Sequence 450, App
Sequence 1930, Ap
Sequence 187302,
Sequence 164243,
Sequence 246428,
Sequence 13693, A
Sequence 197783,
Sequence 165206,
Sequence 246430,

ALIGNMENTS

RESULT 1
US-09-816-277-60
; Sequence 60, Application US/09816277
; Publication No. US20010034045A1
; GENERAL INFORMATION:
; APPLICANT: Penttila, Merja E.
; APPLICANT: Ward, Michael
; APPLICANT: Wang, Huaming
; APPLICANT: Valkonen, Mari J.
; APPLICANT: Saloheimo, Markku
; TITLE OF INVENTION: Increased Production of Secreted
; TITLE OF INVENTION: Proteins by Recombinant Eukaryotic Cells
; FILE REFERENCE: GC590-2
; CURRENT APPLICATION NUMBER: US/09/816, 277
; CURRENT FILING DATE: 2001-03-23
; PRIOR APPLICATION NUMBER: US 09/534, 692
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 60
; LENGTH: 68
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-09-816-277-60

Query Match 100.0%; Score 342; DB 3; Length 68;
Best Local Similarity 100.0%; Pred.No. 2.2e-27;
Matches 68; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KSTLPPRKAKTKEKEQRRERILNRRRAAHOSSREKRLHQLYLRKCSLLENLNSVN 60
|||||
Db 1 KSTLPPRKAKTKEKEQRRERILNRRRAAHOSSREKRLHQLYLRKCSLLENLNSVN 60
|||||

QY 61 LEKLADHE 68
|||||
Db 61 LEKLADHE 68

RESULT 2
US-10-663-450-60
; Sequence 60, Application US/10663450
; Publication No. US20040186070A1
; GENERAL INFORMATION:
; APPLICANT: Penttila, Merja E.
; APPLICANT: Ward, Michael
; APPLICANT: Wang, Huaming
; APPLICANT: Valkonen, Mari J.
; APPLICANT: Saloheimo, Markku
; TITLE OF INVENTION: Increased Production of Secreted
; TITLE OF INVENTION: Proteins by Recombinant Eukaryotic Cells

```
; FILE REFERENCE: GC590-2
; CURRENT APPLICATION NUMBER: US/10/663,450
; PRIOR FILING DATE: 2003-09-15
; PRIOR APPLICATION NUMBER: US 09/534,692
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 60
; LENGTH: 68
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-10-663-450-60

Query Match      100.0%; Score 342; DB 4; Length 68;
Best Local Similarity 100.0%; Pred. No. 2.2e-27;
Matches 68; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KSTLPPRKRAKTKKEEKQRIERILNRRRAAHQSREKKRLHLQYLERKCSLLENLNSVN 60
   |||||
Db 1 KSTLPPRKRAKTKKEEKQRIERILNRRRAAHQSREKKRLHLQYLERKCSLLENLNSVN 60
   |||||
QY 61 LEKLADHE 68
   |||||
Db 61 LEKLADHE 68
   |||||

RESULT 3
US-10-369-493-21867
; Sequence 21867, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 21867
; LENGTH: 200
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
US-10-369-493-21867

Query Match      100.0%; Score 342; DB 4; Length 200;
Best Local Similarity 100.0%; Pred. No. 6.7e-27;
Matches 68; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KSTLPPRKRAKTKKEEKQRIERILNRRRAAHQSREKKRLHLQYLERKCSLLENLNSVN 60
   |||||
Db 23 KSTLPPRKRAKTKKEEKQRIERILNRRRAAHQSREKKRLHLQYLERKCSLLENLNSVN 82
   |||||
QY 61 LEKLADHE 68
   |||||
Db 83 LEKLADHE 90
   |||||

RESULT 4
US-09-816-277-2
; Sequence 2, Application US/09816277
; Publication No. US20010034045A1
; GENERAL INFORMATION:
; APPLICANT: Penttila, Merja E.
; APPLICANT: Ward, Michael
; APPLICANT: Wang, Huaming
; APPLICANT: Valkonen, Mari J.
; APPLICANT: Saloheimo, Markku
```

```
; TITLE OF INVENTION: Increased Production of Secreted
; FILE OF INVENTION: Proteins by Recombinant Eukaryotic Cells
; FILE REFERENCE: GC590-2
; CURRENT APPLICATION NUMBER: US/09/816,277
; CURRENT FILING DATE: 2001-03-23
; PRIOR APPLICATION NUMBER: US 09/534,692
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 451
; TYPE: PRT
; ORGANISM: Trichoderma reesei
US-09-816-277-2

Query Match      57.3%; Score 196; DB 3; Length 451;
Best Local Similarity 61.9%; Pred. No. 1.1e-11;
Matches 39; Conservative 13; Mismatches 9; Indels 2; Gaps 1;

QY 1 KSTLPPRKRAKTKKEEKQRIERILNRRRAAHQSREKKRLHLQYLERKCSLLENLNSVN 60
   |||||
Db 102 KTNLPPRKRAKTEDEKQRRVERVLNRRRAAQSRRERKLEVEALEKRNKELETL--IN 159
   |||||
QY 61 LEK 63
   ||
Db 160 VQK 162
   ||

RESULT 5
US-09-816-277-5
; Sequence 5, Application US/09816277
; Publication No. US20010034045A1
; GENERAL INFORMATION:
; APPLICANT: Penttila, Merja E.
; APPLICANT: Ward, Michael
; APPLICANT: Wang, Huaming
; APPLICANT: Valkonen, Mari J.
; APPLICANT: Saloheimo, Markku
; TITLE OF INVENTION: Increased Production of Secreted
; FILE REFERENCE: GC590-2
; CURRENT APPLICATION NUMBER: US/09/816,277
; CURRENT FILING DATE: 2001-03-23
; PRIOR APPLICATION NUMBER: US 09/534,692
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 451
; TYPE: PRT
; ORGANISM: Trichoderma reesei
US-09-816-277-5

Query Match      57.3%; Score 196; DB 3; Length 451;
Best Local Similarity 61.9%; Pred. No. 1.1e-11;
Matches 39; Conservative 13; Mismatches 9; Indels 2; Gaps 1;

QY 1 KSTLPPRKRAKTKKEEKQRIERILNRRRAAHQSREKKRLHLQYLERKCSLLENLNSVN 60
   |||||
Db 102 KTNLPPRKRAKTEDEKQRRVERVLNRRRAAQSRRERKLEVEALEKRNKELETL--IN 159
   |||||
QY 61 LEK 63
   ||
Db 160 VQK 162
   ||

RESULT 6
US-10-663-450-2
; Sequence 2, Application US/10663450
; Publication No. US20040186070A1
; GENERAL INFORMATION:
; APPLICANT: Penttila, Merja E.
; APPLICANT: Ward, Michael
```

```
; APPLICANT: Wang, Huaming
; APPLICANT: Valkonen, Mari J.
; APPLICANT: Saloheimo, Markku
; TITLE OF INVENTION: Increased Production of Secreted
; FILE REFERENCE: GC590-2
; CURRENT APPLICATION NUMBER: US/10/663,450
; CURRENT FILING DATE: 2003-09-15
; PRIOR APPLICATION NUMBER: US 09/534,692
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 451
; TYPE: PRT
; ORGANISM: Trichoderma reesei
US-10-663-450-2

Query Match      57.3%; Score 196; DB 4; Length 451;
Best Local Similarity 61.9%; Pred. No. 1.1e-11;
Matches 39; Conservative 13; Mismatches 9; Indels 2; Gaps 1;

Qy 1 KSTLPPRKRAKTEKEQRRIILNRRAAHSREKKRLHLQYLERKCSLLENLNSVN 60
   |:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 102 KTNLPKRAKTEDEKQRRVRLNRRAAQSRRERKRLEVEALEKRNKELETLI--IN 159
   |:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

Qy 61 LEK 63
   :||
Db 160 VQK 162

RESULT 7
US-10-663-450-5
; Sequence 5, Application US/10663450
; Publication No. US20040186070A1
; GENERAL INFORMATION:
; APPLICANT: Penttila, Merja E.
; APPLICANT: Ward, Michael
; APPLICANT: Wang, Huaming
; APPLICANT: Valkonen, Mari J.
; APPLICANT: Saloheimo, Markku
; TITLE OF INVENTION: Increased Production of Secreted
; FILE REFERENCE: GC590-2
; CURRENT APPLICATION NUMBER: US/10/663,450
; CURRENT FILING DATE: 2003-09-15
; PRIOR APPLICATION NUMBER: US 09/534,692
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 451
; TYPE: PRT
; ORGANISM: Trichoderma reesei
US-10-663-450-5

Query Match      57.3%; Score 196; DB 4; Length 451;
Best Local Similarity 61.9%; Pred. No. 1.1e-11;
Matches 39; Conservative 13; Mismatches 9; Indels 2; Gaps 1;

Qy 1 KSTLPPRKRAKTEKEQRRIILNRRAAHSREKKRLHLQYLERKCSLLENLNSVN 60
   |:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 102 KTNLPKRAKTEDEKQRRVRLNRRAAQSRRERKRLEVEALEKRNKELETLI--IN 159
   |:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

Qy 61 LEK 63
   :||
Db 160 VQK 162

RESULT 8
US-09-533-029-104
; Sequence 104, Application US/09533029
; Publication No. US20030046723A1
```

```
; GENERAL INFORMATION:
; APPLICANT: Heard, Jacqueline
; APPLICANT: Broun, Pierre
; APPLICANT: Riechmann, Jose-Luis
; APPLICANT: Keddle, James
; APPLICANT: Pineda, Omaira
; APPLICANT: Adam, Luc
; APPLICANT: Samaha, Raymond
; APPLICANT: Zhang, James
; APPLICANT: Yu, Guo-Liang
; APPLICANT: Ratcliffe, Oliver
; APPLICANT: Pilgrim, Mareha
; APPLICANT: Jiang, Cai-Zhong
; APPLICANT: Reuber, Lynne
; TITLE OF INVENTION: DISEASE-INDUCED POLYNUCLEOTIDES
; FILE REFERENCE: MBI-010
; CURRENT APPLICATION NUMBER: US/09/533,029
; CURRENT FILING DATE: 2000-03-22
; EARLIER APPLICATION NUMBER: 60/125,814
; EARLIER FILING DATE: 1999-03-23
; NUMBER OF SEQ ID NOS: 121
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 104
; LENGTH: 409
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; OTHER INFORMATION: G1034
US-09-533-029-104

Query Match      52.2%; Score 178.5; DB 3; Length 409;
Best Local Similarity 56.5%; Pred. No. 5.0e-10;
Matches 39; Conservative 14; Mismatches 13; Indels 3; Gaps 2;

Qy 1 KSTLPPRKRAKTEKEQRRIILNRRAAHSREKKRLHLQYLERKCSLLENLNSVN 60
   |:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 84 KTNLPKRAKTEDEKQRRVRLNRRAAQSRRERKRLEVEALEKRAIER--KNMD 141
   |:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

Qy 61 LE-KLADHE 68
   ||:||||
Db 142 LEMRLADME 150

RESULT 9
US-10-295-403-158
; Sequence 158, Application US/10295403
; Publication No. US20030101481A1
; GENERAL INFORMATION:
; APPLICANT: Heard, Jacqueline
; APPLICANT: Riechmann, Jose Luis
; APPLICANT: Adam, Luc
; APPLICANT: Broun, Pierre
; APPLICANT: Pineda, Omaira
; APPLICANT: Reuber, Lynne
; APPLICANT: Jiang, Cai-Zhong
; APPLICANT: Keddle, James
; APPLICANT: Zhang, James
; APPLICANT: Benito, Maria-Ines
; APPLICANT: Yu, Guo-Liang
; APPLICANT: Fromm, Mike
; TITLE OF INVENTION: PLANT GENE SEQUENCES I
; FILE REFERENCE: MBI-0003
; CURRENT APPLICATION NUMBER: US/10/295,403
; CURRENT FILING DATE: 2002-11-15
; PRIOR APPLICATION NUMBER: US/09/394,519
; PRIOR FILING DATE: 1999-09-13
; PRIOR APPLICATION NUMBER: 60/101,349
; PRIOR FILING DATE: 1998-09-22
; PRIOR APPLICATION NUMBER: 60/103,312
; PRIOR FILING DATE: 1998-10-06
; PRIOR APPLICATION NUMBER: 60/108,734
; PRIOR FILING DATE: 1998-11-17
; PRIOR APPLICATION NUMBER: 60/113,409
```

```
; PRIOR FILING DATE: 1998-12-22
; NUMBER OF SEQ ID NOS: 170
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 158
; LENGTH: 409
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; OTHER INFORMATION: G1034
US-10-295-403-158

Query Match          52.2%; Score 178.5; DB 4; Length 409;
Best Local Similarity 56.5%; Pred. No. 5.8e-10;
Matches 39; Conservative 14; Mismatches 13; Indels 3; Gaps 2;

QY 1 KSTLPKRAKTKKEKQRRRIILNRRRAAHQSREKKRLHLQYLERKCSLLENLNSYN 60
   |:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 84 KTNLPKRAKTKDEKQRRVERVLNRRRAAQSRRKRKEVEALEVEKRAIER--KNMD 141
   |:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 61 LE-KLADHE 68
   ||:|||||
Db 142 LEMRLADME 150

RESULT 10
US-10-412-699B-684
; Sequence 684, Application US/10412699B
; Publication No. US20040045049A1
; GENERAL INFORMATION:
; APPLICANT: Mendel Biotechnology, Inc.
; APPLICANT: Zhang, James
; APPLICANT: Fromm, Michael E.
; APPLICANT: Heard, Jacqueline E.
; APPLICANT: Riechmann, Jose Luis
; APPLICANT: Adam, Luc J.
; APPLICANT: Broun, Pierre E.
; APPLICANT: Pineda, Omaira
; APPLICANT: Reuber, T. Lynne
; APPLICANT: Keddle, James S.
; APPLICANT: Yu, Guo-Liang
; APPLICANT: Jiang, Cai-Zhong
; APPLICANT: Samaha, Raymond R.
; APPLICANT: Pilgrim, Marsha L.
; APPLICANT: Creelman, Robert A.
; APPLICANT: DuBell, Arnold N.
; APPLICANT: Ratcliffe, Oliver
; APPLICANT: Kumamoto, Roderick
; APPLICANT: Sherman, Bradley K.
; TITLE OF INVENTION: Polynucleotides and Polypeptides in Plants
; FILE REFERENCE: MBI-0048CIP
; CURRENT APPLICATION NUMBER: US/10/412,699B
; CURRENT FILING DATE: 2003-04-10
; PRIOR APPLICATION NUMBER: 09/394,519
; PRIOR FILING DATE: 1999-09-13
; PRIOR APPLICATION NUMBER: 09/489,376
; PRIOR FILING DATE: 2000-01-21
; PRIOR APPLICATION NUMBER: 09/506,720
; PRIOR FILING DATE: 2000-02-17
; PRIOR APPLICATION NUMBER: 09/533,030
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: 09/533,392
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: 09/533,029
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: 09/532,591
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: 09/533,648
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: 09/713,994
; PRIOR FILING DATE: 2000-11-16
; PRIOR APPLICATION NUMBER: 09/819,142
; PRIOR FILING DATE: 2001-03-27
; Remaining Prior Application data removed - See File Wrapper or PALM.

; NUMBER OF SEQ ID NOS: 2011
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 684
; LENGTH: 409
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; OTHER INFORMATION: G1034
US-10-412-699B-684

Query Match          52.2%; Score 178.5; DB 4; Length 409;
Best Local Similarity 56.5%; Pred. No. 5.8e-10;
Matches 39; Conservative 14; Mismatches 13; Indels 3; Gaps 2;

QY 1 KSTLPKRAKTKKEKQRRRIILNRRRAAHQSREKKRLHLQYLERKCSLLENLNSYN 60
   |:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 84 KTNLPKRAKTKDEKQRRVERVLNRRRAAQSRRKRKEVEALEVEKRAIER--KNMD 141
   |:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 61 LE-KLADHE 68
   ||:|||||
Db 142 LEMRLADME 150

RESULT 11
US-09-816-277-19
; Sequence 19, Application US/09816277
; Publication No. US20010034045A1
; GENERAL INFORMATION:
; APPLICANT: Penttila, Merja E.
; APPLICANT: Ward, Michael
; APPLICANT: Wang, Huaming
; APPLICANT: Valkonen, Mari J.
; APPLICANT: Saloheimo, Markku
; TITLE OF INVENTION: Increased Production of Secreted
; TITLE OF INVENTION: Proteins by Recombinant Eukaryotic Cells
; FILE REFERENCE: GCS90-2
; CURRENT APPLICATION NUMBER: US/09/816,277
; CURRENT FILING DATE: 2001-03-23
; PRIOR APPLICATION NUMBER: US 09/534,692
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 19
; LENGTH: 342
; TYPE: PRT
; ORGANISM: Aspergillus niger
US-09-816-277-19

Query Match          51.6%; Score 176.5; DB 3; Length 342;
Best Local Similarity 55.9%; Pred. No. 7.7e-10;
Matches 38; Conservative 12; Mismatches 17; Indels 1; Gaps 1;

QY 1 KSTLPKRAKTKKEKQRRRIILNRRRAAHQSREKKRLHLQYLERKCSLLENLNSYN 60
   |:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 64 KTNLPKRAKTKDEKQRRRIILNRRRAAQSRRKRKEVEALEVEKRAIER--QNQFL 122
   |:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 61 LEKLADHE 68
   |:|||||
Db 123 LQRLSQME 130

RESULT 12
US-10-663-450-19
; Sequence 19, Application US/10663450
; Publication No. US20040186070A1
; GENERAL INFORMATION:
; APPLICANT: Penttila, Merja E.
; APPLICANT: Ward, Michael
; APPLICANT: Wang, Huaming
; APPLICANT: Valkonen, Mari J.
; APPLICANT: Saloheimo, Markku
; TITLE OF INVENTION: Increased Production of Secreted
; TITLE OF INVENTION: Proteins by Recombinant Eukaryotic Cells
```



```
; FILE REFERENCE: GC590-2
; CURRENT APPLICATION NUMBER: US/10/663,450
; CURRENT FILING DATE: 2003-09-15
; PRIOR APPLICATION NUMBER: US 09/534,692
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 19
; LENGTH: 342
; TYPE: PRT
; ORGANISM: Aspergillus niger
US-10-663-450-19

Query Match
Best Local Similarity 51.6%; Score 176.5; DB 4; Length 342;
Matches 38; Conservative 12; Mismatches 17; Indels 1; Gaps 1;

QY 1 KSTLPPRKAKTKKEQRRRIILNRRAAHQSRKKRLHLYLQYLRKCSLLENLNSVN 60
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 64 KTNLPKRKAKTEDEKQRRRIERVLNRRAAAQTSRKRKLEMEKLESEKIDMEQ-QNQFL 122
   :|::|:|

QY 61 LEKLADHE 68
   :|::|:|
Db 123 LQRLSQME 130

RESULT 13
US-09-816-277-4
; Sequence 4, Application US/09816277
; Publication No. US20010034045A1
; GENERAL INFORMATION:
; APPLICANT: Penttila, Merja E.
; APPLICANT: Ward, Michael
; APPLICANT: Wang, Huaming
; APPLICANT: Valkonen, Mari J.
; APPLICANT: Saloheimo, Markku
; TITLE OF INVENTION: Increased Production of Secreted
; FILE REFERENCE: GC590-2
; CURRENT APPLICATION NUMBER: US/09/816,277
; CURRENT FILING DATE: 2001-03-23
; PRIOR APPLICATION NUMBER: US 09/534,692
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 349
; TYPE: PRT
; ORGANISM: Aspergillus nidulans
US-09-816-277-4

Query Match
Best Local Similarity 51.6%; Score 176.5; DB 3; Length 349;
Matches 39; Conservative 11; Mismatches 17; Indels 1; Gaps 1;

QY 1 KSTLPPRKAKTKKEQRRRIILNRRAAHQSRKKRLHLYLQYLRKCSLLENLNSVN 60
   :|::|:|
Db 70 KTNLPKRKAKTEDEKQRRRIERVLNRRAAAQTSRKRKLEMEKLESEKIDMEQ-QNQFL 128
   :|::|:|

QY 61 LEKLADHE 68
   :|::|:|
Db 129 LQRLAQME 136

RESULT 14
US-09-816-277-6
; Sequence 6, Application US/09816277
; Publication No. US20010034045A1
; GENERAL INFORMATION:
; APPLICANT: Penttila, Merja E.
; APPLICANT: Ward, Michael
; APPLICANT: Wang, Huaming
; APPLICANT: Valkonen, Mari J.
; TITLE OF INVENTION: Increased Production of Secreted
; FILE REFERENCE: GC590-2
; CURRENT APPLICATION NUMBER: US/09/816,277
; CURRENT FILING DATE: 2003-09-15
; PRIOR APPLICATION NUMBER: US 09/534,692
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 349
; TYPE: PRT
; ORGANISM: Aspergillus nidulans
US-10-663-450-4

Query Match
Best Local Similarity 51.6%; Score 176.5; DB 4; Length 349;
Matches 39; Conservative 11; Mismatches 17; Indels 1; Gaps 1;

QY 1 KSTLPPRKAKTKKEQRRRIILNRRAAHQSRKKRLHLYLQYLRKCSLLENLNSVN 60
   :|::|:|
Db 70 KTNLPKRKAKTEDEKQRRRIERVLNRRAAAQTSRKRKLEMEKLESEKIDMEQ-QNQFL 128
   :|::|:|

QY 61 LEKLADHE 68
   :|::|:|
Db 129 LQRLAQME 136

Search completed: November 23, 2005, 03:32:22
Job time : 23.0138 secs
```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 22, 2005, 15:43:54 ; Search time 4 Seconds
(without alignments)
18.077 Million cell updates/sec

Title: US-10-663-450-6_COPY_53_116

Perfect score: 324

Sequence: 1 KRPAAKRRKSGWQLPVKTN.....RAAAQTSRRKRLEMEKLS 64

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 8323 seqs, 1129788 residues

Total number of hits satisfying chosen parameters: 8323

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : 1: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB pep.*

2: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB pep.*

3: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB pep.*

4: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB pep.*

5: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB pep.*

6: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB pep.*

7: /cgn2_6/ptodata/1/pubpaa/US11_NEW_PUB pep.*

8: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	91	28.1	376	1	US-10-971-994-2
2	69.5	21.5	605	1	US-10-131-826A-160
3	65.5	20.2	431	7	US-11-169-013-2
4	64	19.8	454	7	US-11-169-013-1
5	63.5	19.6	879	7	US-11-022-562-340
6	59	18.2	400	7	US-11-018-018-7
7	56	17.3	134	1	US-10-329-472-2
8	55.5	17.1	194	7	US-11-074-176-82
9	55	17.0	625	1	US-10-510-386-6
10	54.5	16.8	615	1	US-10-982-545-14
11	54	16.7	21	1	US-10-995-951A-33
12	54	16.7	903	1	US-10-689-742-142
13	53	16.4	188	1	US-10-131-826A-2
14	51.5	15.9	860	7	US-11-022-562-217
15	51	15.7	312	1	US-10-986-501-124
16	51	15.7	747	7	US-11-182-592-2
17	51	15.7	804	1	US-10-467-962B-57
18	51	15.7	1302	7	US-11-004-057-6
19	51	15.7	1493	7	US-11-004-057-4
20	51	15.7	1493	7	US-11-004-057-21
21	50.5	15.6	128	7	US-11-105-864-4
22	50.5	15.6	240	1	US-10-689-742-210
23	50.5	15.6	1618	1	US-10-984-645-2
24	49.5	15.3	718	1	US-10-467-962B-29
25	49.5	15.3	718	1	US-10-467-962B-65

26	49	15.1	1023	1	US-10-131-826A-200	Sequence 200, App
27	48.5	15.0	180	1	US-10-689-742-84	Sequence 84, Appl
28	48.5	15.0	327	7	US-11-090-878-16	Sequence 16, Appl
29	48	14.8	2897	1	US-10-499-715-2	Sequence 2, Appli
30	47.5	14.7	234	1	US-10-997-697-47	Sequence 47, Appl
31	47.5	14.7	453	1	US-10-131-826A-270	Sequence 270, App
32	47.5	14.7	476	1	US-10-997-697-29	Sequence 29, Appl
33	47.5	14.7	526	1	US-10-997-697-28	Sequence 28, Appl
34	47.5	14.7	790	1	US-10-131-826A-204	Sequence 204, App
35	47.5	14.7	1267	7	US-11-109-156-35	Sequence 35, Appl
36	47	14.5	127	1	US-10-997-437A-8	Sequence 8, Appli
37	47	14.5	221	1	US-10-636-716-60	Sequence 60, Appl
38	47	14.5	877	7	US-11-074-176-322	Sequence 322, App
39	47	14.5	883	7	US-11-074-176-88	Sequence 88, Appl
40	46.5	14.4	26	7	US-11-019-894A-20	Sequence 20, Appl
41	46.5	14.4	197	1	US-10-986-501-296	Sequence 296, App
42	46.5	14.4	435	1	US-10-510-386-62	Sequence 62, Appl
43	46.5	14.4	586	1	US-10-131-826A-46	Sequence 46, Appl
44	46	14.2	616	1	US-10-982-545-5	Sequence 5, Appli
45	46	14.2	683	7	US-11-046-668-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1

US-10-971-994-2

; Sequence 2, Application US/10971994

; Publication No. US20050250182A1

; GENERAL INFORMATION:

; APPLICANT: University of Michigan et al.

; APPLICANT: Kaufman, Randal

; APPLICANT: Kyungso, Lee

; APPLICANT: Mori, Kazutos

; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, AND METHODS FOR

; FILE REFERENCE: MODULATING THE UNFOLDED PROTEIN RESPONSE

; FILE REFERENCE: UMV-2246

; CURRENT APPLICATION NUMBER: US/10/971,994

; CURRENT FILING DATE: 2004-10-21

; PRIOR APPLICATION NUMBER: PCT/2003/012640

; PRIOR FILING DATE: 2003-04-22

; PRIOR APPLICATION NUMBER: 60/375,098

; PRIOR FILING DATE: 2002-04-22

; PRIOR APPLICATION NUMBER: 60/374,880

; PRIOR FILING DATE: 2002-04-23

; NUMBER OF SEQ ID NOS: 21

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 2

; LENGTH: 376

; TYPE: PRT

; ORGANISM: Homo sapiens

US-10-971-994-2

Query Match 28.1%; Score 91; DB 1; Length 376;

Best Local Similarity 40.3%; Pred. No. 0.00045;

Matches 27; Conservative 9; Mismatches 17; Indels 14; Gaps 4;

QY 11 GQELP--VP-----KTNLP-PRKRAK-TEDEKEQRRIERVLRNAAQTSRRKR 56

Db 33 GQALPLMVPAQRGASPEASGGLPQARKQRLTHLSPEEKALRRKLKRVAAQTARDKK 92

QY 57 LEMEKLE 63

Db 93 ARMSELE 99

RESULT 2

US-10-131-826A-160

; Sequence 160, Application US/10131826A

; Publication No. US20050245730A1

; GENERAL INFORMATION:

; APPLICANT: Baker, Kevin P.

; APPLICANT: Beresini, Maureen

```
/ APPLICANT: DeForge,Laura
/ APPLICANT: Desnoyers,Luc
/ APPLICANT: Filvaroff,Ellen
/ APPLICANT: Gao,Wei-Qiang
/ APPLICANT: Gerritsen,Mary E.
/ APPLICANT: Goddard,Audrey
/ APPLICANT: Godowski,Paul J.
/ APPLICANT: Gurney,Austin L.
/ APPLICANT: Sherwood,Steven
/ APPLICANT: Smith,Victoria
/ APPLICANT: Stewart,Timothy A.
/ APPLICANT: Tumas,Daniel
/ APPLICANT: Watanabe,Colin K
/ APPLICANT: Wood,William
/ APPLICANT: Zhang, Zemin
/ TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
/ FILE REFERENCE: P3330R1C128
/ CURRENT APPLICATION NUMBER: US/10/131,826A
/ CURRENT FILING DATE: 2002-04-24
/ PRIOR APPLICATION NUMBER: 60/049911
/ PRIOR FILING DATE: 1997-06-18
/ PRIOR APPLICATION NUMBER: 60/056974
/ PRIOR FILING DATE: 1997-08-26
/ PRIOR APPLICATION NUMBER: 60/059113
/ PRIOR FILING DATE: 1997-09-17
/ PRIOR APPLICATION NUMBER: 60/059115
/ PRIOR FILING DATE: 1997-09-17
/ PRIOR APPLICATION NUMBER: 60/059117
/ PRIOR FILING DATE: 1997-09-17
/ PRIOR APPLICATION NUMBER: 60/059122
/ PRIOR FILING DATE: 1997-09-17
/ PRIOR APPLICATION NUMBER: 60/059184
/ PRIOR FILING DATE: 1997-09-17
/ PRIOR APPLICATION NUMBER: 60/059263
/ PRIOR FILING DATE: 1997-09-18
/ PRIOR APPLICATION NUMBER: 60/059352
/ PRIOR FILING DATE: 1997-09-19
/ PRIOR APPLICATION NUMBER: 60/059588
/ PRIOR FILING DATE: 1997-09-19
/ Remaining Prior Application data removed - See File Wrapper or PALM.
/ NUMBER OF SEQ ID NOS: 550
/ SEQ ID NO 160
/ LENGTH: 605
/ TYPE: PRT
/ ORGANISM: Homo Sapien
US-10-131-826A-160

Query Match          21.5%; Score 69.5; DB 1; Length 605;
Best Local Similarity 34.4%; Pred. No. 0.18;
Matches 21; Conservative 8; Mismatches 15; Indels 17; Gaps 3;

QY 9 SNGQEL-----PVPKTNL-----PPRKRAKTEDEKEORRIERVLRNRAAAQTSRERKR 56
DB 134 SQAQLPFLPPLPKMNLVPPPHMPPRPEEEEEERKEEVEK-----QEEEEEE 188
QY 57 L 57
DB 189 L 189

RESULT 3
US-11-169-013-2
/ Sequence 2, Application US/11169013
/ Publication No. US20050244971A1
/ GENERAL INFORMATION:
/ APPLICANT: Korea Kumho Petrochemical Co., Ltd.
/ TITLE OF INVENTION: TRANSGENIC PLANTS WITH ENHANCED STRESS TOLERANCE
/ FILE REFERENCE:
/ CURRENT APPLICATION NUMBER: US/11/169,013
/ CURRENT FILING DATE: 2005-06-29
/ NUMBER OF SEQ ID NOS: 34
/ SOFTWARE: Kopatentin 1.71
/ SEQ ID NO 1
/ LENGTH: 171
/ TYPE: PRT
/ ORGANISM: Simian Immunodeficiency Virus
US-11-022-562-340

Query Match          19.6%; Score 63.5; DB 7; Length 879;
Best Local Similarity 28.0%; Pred. No. 1.2;
Matches 21; Conservative 14; Mismatches 14; Indels 13; Gaps 3;

QY 2 KPAKPKSNG-----QELPVPKTNLPPrKRAKTEDEKEORRIERVLRNRAA-----AQ 49
DB 601 KDAQLNAGCAFRQVCHTTPVPNTSLTPKWDNETWQWE-RKVDFLSENLALPEEAQ 659
```

```
/ SEQ ID NO 2
/ LENGTH: 431
/ TYPE: PRT
/ ORGANISM: Arabidopsis thaliana
US-11-169-013-2

Query Match          20.2%; Score 65.5; DB 7; Length 431;
Best Local Similarity 38.6%; Pred. No. 0.34;
Matches 22; Conservative 10; Mismatches 16; Indels 9; Gaps 3;

QY 15 PVPKTNLPPrKRAKTEDEKEORRIERVLRNRAAAQTSRERKR-----LEMEKLE 63
DB 332 PVPYV-LNRGRSNTGLEKVIERRQRMKKNRESAARSARKQAYTLELEAEIKLK 387

RESULT 4
US-11-169-013-1
/ Sequence 1, Application US/11169013
/ Publication No. US20050244971A1
/ GENERAL INFORMATION:
/ APPLICANT: Korea Kumho Petrochemical Co., Ltd.
/ TITLE OF INVENTION: TRANSGENIC PLANTS WITH ENHANCED STRESS TOLERANCE
/ FILE REFERENCE:
/ CURRENT APPLICATION NUMBER: US/11/169,013
/ CURRENT FILING DATE: 2005-06-29
/ NUMBER OF SEQ ID NOS: 34
/ SOFTWARE: Kopatentin 1.71
/ SEQ ID NO 1
/ LENGTH: 454
/ TYPE: PRT
/ ORGANISM: Arabidopsis thaliana
US-11-169-013-1

Query Match          19.8%; Score 64; DB 7; Length 454;
Best Local Similarity 34.0%; Pred. No. 0.52;
Matches 17; Conservative 11; Mismatches 22; Indels 0; Gaps 0;

QY 15 PVPKTNLPPrKRAKTEDEKEORRIERVLRNRAAAQTSRERKRLEMEKLES 64
DB 353 PVPYMGVRKGTGAVLEKVIERRQRMKKNRESAARSARKQAYTMELEA 402
```

```
RESULT 5
US-11-022-562-340
/ Sequence 340, Application US/11022562
/ Publication No. US20050249742A1
/ GENERAL INFORMATION:
/ APPLICANT: Ruprecht, Ruth M.
/ APPLICANT: Shisong, Jiang
/ TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR MODULATING
/ TITLE OF INVENTION: A CYTOTOXIC T LYMPHOCYTE IMMUNE RESPONSE
/ FILE REFERENCE: DFN-043CN
/ CURRENT APPLICATION NUMBER: US/11/022,562
/ CURRENT FILING DATE: 2004-12-22
/ PRIOR APPLICATION NUMBER: PCT/US03/20322
/ PRIOR FILING DATE: 2003-06-27
/ PRIOR APPLICATION NUMBER: 60/392718
/ PRIOR FILING DATE: 2002-06-27
/ NUMBER OF SEQ ID NOS: 340
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 340
/ LENGTH: 879
/ TYPE: PRT
/ ORGANISM: Simian Immunodeficiency Virus
US-11-022-562-340
```

```
Query Match          19.6%; Score 63.5; DB 7; Length 879;
Best Local Similarity 28.0%; Pred. No. 1.2;
Matches 21; Conservative 14; Mismatches 14; Indels 13; Gaps 3;

QY 2 KPAKPKSNG-----QELPVPKTNLPPrKRAKTEDEKEORRIERVLRNRAA-----AQ 49
DB 601 KDAQLNAGCAFRQVCHTTPVPNTSLTPKWDNETWQWE-RKVDFLSENLALPEEAQ 659
```

```

Qy   50 TSRRKRLEMEKLES 64
      :|: ||::|| |
Db    660 IQQEKMYEQKLNS 674

RESULT 6
US-11-018-018-7
; Sequence 7, Application US/11018018
; Publication No. US20050250794A1
; GENERAL INFORMATION:
; APPLICANT: Napper, Andrew
; APPLICANT: Distefano, Peter
; APPLICANT: Curtis, Rory
; APPLICANT: Mcdonagh, Thomas
; APPLICANT: Hixon, Jeffrey
; TITLE OF INVENTION: METHODS OF TREATING A DISORDER
; FILE REFERENCE: 13407-052001
; CURRENT APPLICATION NUMBER: US/11/018,018
; CURRENT FILING DATE: 2004-12-20
; PRIOR APPLICATION NUMBER: US 60/530,945
; PRIOR FILING DATE: 2003-12-19
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 400
; TYPE: PRT
; ORGANISM: Homo sapiens
US-11-018-018-7

Query Match          18.2%; Score 59; DB 7; Length 400;
Best Local Similarity 37.5%; Pred. No. 1.6;
Matches 15; Conservative 11; Mismatches 12; Indels 2; Gaps 2;

Qy   25 KRAKTEDEKEQ-RRIERVLRNAAQAOTSRERKRLEMEKLE 63
      :|: ||::|| |
Db    18 RRLREEQQRRLRVQSRIIR-KAAAERSAEGRLLAESAD 56

RESULT 7
US-10-329-472-2
; Sequence 2, Application US/10329472
; Publication No. US20050244888A1
; GENERAL INFORMATION:
; APPLICANT: Li et al.
; TITLE OF INVENTION: Human Chemokine Beta-9
; FILE REFERENCE: PF131P2D1
; CURRENT APPLICATION NUMBER: US/10/329,472
; CURRENT FILING DATE: 2002-12-27
; PRIOR APPLICATION NUMBER: US 08/793,381
; PRIOR FILING DATE: 1997-05-19
; PRIOR APPLICATION NUMBER: PCI/US95/06260
; PRIOR FILING DATE: 1996-02-29
; PRIOR APPLICATION NUMBER: US 08/294,251
; PRIOR FILING DATE: 1994-08-23
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 134
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-329-472-2

Query Match          17.3%; Score 56; DB 1; Length 134;
Best Local Similarity 25.3%; Pred. No. 1;
Matches 20; Conservative 15; Mismatches 20; Indels 24; Gaps 3;

Qy   1  KKPAKKRKSWGQELP-----VPKTNLPPRKAKTE---DEKE-----QR 36
      |||||:::|::|:|
Db   39 KIPAKVRSYRKQPSLGCSSIPAILFLFPRKKSQAELCADPKELWVQLMQHLDKTPSPQK 98
      |||||:::|::|:|
Qy   37 RIERVLRNAAQAOTSRERK 55
      :|: ||::|| |
```

```
US-10-982-545-14
; Sequence 14, Application US/10982545
; Publication No. US20050244890A1
; GENERAL INFORMATION:
; APPLICANT: Davies, Huw Alun
; APPLICANT: McGuire, James
; APPLICANT: Simonsen, Anja Hviid
; APPLICANT: Blennow, Kaj
; APPLICANT: Podust, Vladimir
; APPLICANT: Ciplergen Biosystems, Inc.
; TITLE OF INVENTION: Biomarkers for Alzheimer's Disease
; FILE REFERENCE: 016866-011550US
; CURRENT APPLICATION NUMBER: US/10/982,545
; CURRENT FILING DATE: 2004-11-06
; PRIOR APPLICATION NUMBER: US 60/518,360
; PRIOR FILING DATE: 2003-11-07
; PRIOR APPLICATION NUMBER: US 60/526,753
; PRIOR FILING DATE: 2003-12-02
; PRIOR APPLICATION NUMBER: US 60/546,423
; PRIOR FILING DATE: 2004-02-19
; PRIOR APPLICATION NUMBER: US 60/547,250
; PRIOR FILING DATE: 2004-02-23
; PRIOR APPLICATION NUMBER: US 60/558,896
; PRIOR FILING DATE: 2004-04-02
; PRIOR APPLICATION NUMBER: US 60/572,617
; PRIOR FILING DATE: 2004-05-18
; PRIOR APPLICATION NUMBER: US 60/586,503
; PRIOR FILING DATE: 2004-07-08
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 14
; LENGTH: 615
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: nerve growth factor inducible Neurosecretory
; OTHER INFORMATION: protein vaccinia virus growth factor (VGF)
; OTHER INFORMATION: precursor
; NAME/KEY: SIGNAL
; LOCATION: (1)..(22)
; OTHER INFORMATION: signal peptide
; FEATURE:
; NAME/KEY: PEPTIDE
; LOCATION: (23)..(615)
; OTHER INFORMATION: nerve growth factor inducible Neurosecretory
; OTHER INFORMATION: protein vaccinia virus growth factor (VGF)
; FEATURE:
; NAME/KEY: PEPTIDE
; LOCATION: (373)..(417)
; OTHER INFORMATION: biomarker peptide 4808 Da, 4812 Da (Q10),
; OTHER INFORMATION: processed fragment of nerve growth factor
; OTHER INFORMATION: inducible Neurosecretory protein vaccinia virus
; OTHER INFORMATION: growth factor (VGF)
US-10-982-545-14
Query Match 16.8%; Score 54.5; DB 1; Length 615;
Best Local Similarity 27.7%; Pred. No. 8.1;
Matches 18; Conservative 11; Mismatches 13; Indels 23; Gaps 3;

QY 4 AKKKRKGWQELPVPKTNLPPTPR----KRAKTEDEKEQRIIVRLNRRAAAQTSRRKRRLSM 59
Db 560 ALRRRHHYHALP-PSRHYPGREAAQARRAQEEAAEERRLQ-----EQ 600
QY 60 EKLES 64
Db 601 EELEN 605

RESULT 11
US-10-995-951A-33
; Sequence 33, Application US/10995951A
; Publication No. US20050245732A1
```

```
; GENERAL INFORMATION:
; APPLICANT: Hannoufa, A. et al.
; TITLE OF INVENTION: A Repressor-Mediated Regulation System for Control of Gene Expression
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 1096.021B
; CURRENT APPLICATION NUMBER: US/10/995,951A
; CURRENT FILING DATE: 2004-11-23
; PRIOR APPLICATION NUMBER: PCT/CA02/01807
; PRIOR FILING DATE: 2002-11-21
; PRIOR APPLICATION NUMBER: PCT/CA02/00740
; PRIOR FILING DATE: 2002-05-23
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 33
; LENGTH: 21
; TYPE: PRT
; ORGANISM: Nicotiana sp.
US-10-995-951A-33
Query Match 16.7%; Score 54; DB 1; Length 21;
Best Local Similarity 58.8%; Pred. No. 0.23;
Matches 10; Conservative 5; Mismatches 2; Indels 0; Gaps 0;

QY 40 RVLNRRAAAQTSRRKR 56
Db 5 RLVNRRESAQLSRQRKK 21

RESULT 12
US-10-689-742-142
; Sequence 142, Application US/10689742
; Publication No. US20050250180A1
; GENERAL INFORMATION:
; APPLICANT: Jacobs, Kenneth
; APPLICANT: McCoy, John M
; APPLICANT: Lavallie, Edward R
; APPLICANT: Racie, Lisa A
; APPLICANT: Evans, Cheryl
; APPLICANT: Merberg, David
; APPLICANT: Treacy, Maurice
; APPLICANT: Spaulding, Vikki
; TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES ENCODING THEM
; FILE REFERENCE: 00766.000091.10
; CURRENT APPLICATION NUMBER: US/10/689,742
; CURRENT FILING DATE: 2003-10-22
; PRIOR APPLICATION NUMBER: 09/746,783
; PRIOR FILING DATE: 2000-12-21
; NUMBER OF SEQ ID NOS: 231
; SOFTWARE: Patentin version 3.2
; SEQ ID NO 142
; LENGTH: 903
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-689-742-142
Query Match 16.7%; Score 54; DB 1; Length 903;
Best Local Similarity 23.5%; Pred. No. 14;
Matches 19; Conservative 11; Mismatches 19; Indels 32; Gaps 3;

QY 2 KPAKRGKSGWQELPVPKTNLPPTPRKRAKTEDEKEQ-----RRIR 40
Db 72 KQOSTRESW-----VSPRKGLSSEKDNIERQAIENCERRQTEFPVSPVLKRIKR 121
QY 41 VLNRRAAAQTSRRKRLEMEK 61
Db 122 CLRSE-APNSSEEDSPIKSDK 141

RESULT 13
US-10-131-826A-2
; Sequence 2, Application US/10131826A
; Publication No. US20050245730A1
; GENERAL INFORMATION:
```

```
; APPLICANT: Baker, Kevin P.
; APPLICANT: Beresini, Maureen
; APPLICANT: DeGeorge, Laura
; APPLICANT: DeNovoyers, Luc
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Sherwood, Steven
; APPLICANT: Smith, Victoria
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K
; APPLICANT: Wood, William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330R1C128
; CURRENT APPLICATION NUMBER: US/10/131,826A
; CURRENT FILING DATE: 2002-04-24
; PRIOR APPLICATION NUMBER: 60/049911
; PRIOR FILING DATE: 1997-06-18
; PRIOR APPLICATION NUMBER: 60/056974
; PRIOR FILING DATE: 1997-08-26
; PRIOR APPLICATION NUMBER: 60/059113
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059115
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059117
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059122
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059184
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059263
; PRIOR FILING DATE: 1997-09-18
; PRIOR APPLICATION NUMBER: 60/059352
; PRIOR FILING DATE: 1997-09-19
; PRIOR APPLICATION NUMBER: 60/059588
; PRIOR FILING DATE: 1997-09-19
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 2
; LENGTH: 188
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-131-826A-2

Query Match 16.4%; Score 53; DB 1; Length 188;
Best Local Similarity 28.6%; Pred. No. 3.2;
Matches 12; Conservative 13; Mismatches 15; Indels 2; Gaps 1;

QY 25 KRAKTEDEKEQRRIERVLNRRAAAQTSRERKLEM--EKLES 64
DB 37 KKLKAEVQSKKLEKKETITTESAGROQKKKIERQEEKLN 78

RESULT 14
US-11-022-562-217
; Sequence 217, Application US/11022562
; Publication No. US20050249742A1
; GENERAL INFORMATION:
; APPLICANT: Ruprecht, Ruth M.
; APPLICANT: Shisong, Jiang
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR MODULATING
; FILE REFERENCE: DFN-043CN
; CURRENT APPLICATION NUMBER: US/11/022,562
; CURRENT FILING DATE: 2004-12-22
; PRIOR APPLICATION NUMBER: PCT/US03/20322
; PRIOR FILING DATE: 2003-06-27
```

```
; PRIOR APPLICATION NUMBER: 60/392718
; PRIOR FILING DATE: 2002-06-27
; NUMBER OF SEQ ID NOS: 340
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 217
; LENGTH: 860
; TYPE: PRT
; ORGANISM: Human Immunodeficiency Virus 2
US-11-022-562-217

Query Match 15.9%; Score 51.5; DB 7; Length 860;
Best Local Similarity 28.4%; Pred. No. 25;
Matches 19; Conservative 12; Mismatches 25; Indels 11; Gaps 3;

QY 9 SWG-----QELPVPKTNLPKRKRAKT--BDEKEQRRIE-RVLNRRAAAQTSRERKRL 57
DB 593 SNGCAPRQCVCHTTPVWVNDLSLSPDWKNTWQEWKQVRYLEANISQSLEBAQIQEKNMY 652
QY 58 ENEKLES 64
DB 653 ELQKLNS 659

RESULT 15
US-10-986-501-124
; Sequence 124, Application US/10986501
; Publication No. US20050244845A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: 90 Human Secreted Proteins
; FILE REFERENCE: P2013P2C1
; CURRENT APPLICATION NUMBER: US/10/986,501
; CURRENT FILING DATE: 2004-11-12
; PRIOR APPLICATION NUMBER: US/10/621,363
; PRIOR FILING DATE: 2003-07-18
; PRIOR APPLICATION NUMBER: 09/969,730
; PRIOR FILING DATE: 2001-10-06
; PRIOR APPLICATION NUMBER: 09/774,639
; PRIOR FILING DATE: 2001-02-01
; PRIOR APPLICATION NUMBER: 60/238,291
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 09/244,112
; PRIOR FILING DATE: 1999-02-04
; PRIOR APPLICATION NUMBER: PCT/US98/16235
; PRIOR FILING DATE: 1998-08-04
; PRIOR APPLICATION NUMBER: 60/056,371
; PRIOR FILING DATE: 1997-08-19
; PRIOR APPLICATION NUMBER: 60/056,732
; PRIOR FILING DATE: 1997-08-19
; PRIOR APPLICATION NUMBER: 60/056,366
; PRIOR FILING DATE: 1997-08-19
; PRIOR APPLICATION NUMBER: 60/056,364
; PRIOR FILING DATE: 1997-08-19
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 373
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 124
; LENGTH: 312
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-986-501-124

Query Match 15.7%; Score 51; DB 1; Length 312;
Best Local Similarity 25.5%; Pred. No. 9.3;
Matches 12; Conservative 11; Mismatches 22; Indels 2; Gaps 1;
```

```
QY 13 ELPPVPKTNL--PPRKRAKTEDEKEQRRIERVLNRRAAAQTSRERKRL 57
DB 233 KLPEPSASLPNPPSKKIKLSDEPVEAKEDYTKFNTKDLKTEKNSKM 279
```

Search completed: November 22, 2005, 15:54:45
Job time : 4 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 22, 2005, 15:38:17 ; Search time 111 Seconds
(without alignments)
240.911 Million cell updates/sec

Title: US-10-663-450-6_COPY_53_116

Perfect score: 324

Sequence: 1 KKPAAKRSWGQELPVKTN.....RAAAQTSRRKRLMEKLES 64

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications AA Main:*

- 1: /cgn2_6/prodata/1/pubpaa/US07_PUBCOMB.pap:*
- 2: /cgn2_6/prodata/1/pubpaa/US08_PUBCOMB.pap:*
- 3: /cgn2_6/prodata/1/pubpaa/US09_PUBCOMB.pap:*
- 4: /cgn2_6/prodata/1/pubpaa/US10A_PUBCOMB.pap:*
- 5: /cgn2_6/prodata/1/pubpaa/US10B_PUBCOMB.pap:*
- 6: /cgn2_6/prodata/1/pubpaa/US11_PUBCOMB.pap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	324	100.0	349	3	US-09-816-277-4
2	324	100.0	349	3	US-09-816-277-6
3	324	100.0	349	4	US-10-663-450-4
4	324	100.0	349	4	US-10-663-450-6
5	317	97.8	342	3	US-09-816-277-19
6	317	97.8	342	4	US-10-663-450-19
7	317	97.8	386	3	US-09-816-277-16
8	317	97.8	386	4	US-10-663-450-16
9	283	87.3	451	3	US-09-816-277-2
10	283	87.3	451	3	US-09-816-277-5
11	283	87.3	451	4	US-10-663-450-2
12	283	87.3	451	4	US-10-663-450-5
13	277	85.5	409	3	US-09-533-029-104
14	277	85.5	409	4	US-10-295-403-158
15	277	85.5	409	4	US-10-412-699B-684
16	250	77.2	174	4	US-10-767-701-61238
17	171.5	52.9	200	4	US-10-369-493-21867
18	170	52.5	68	3	US-09-816-277-60
19	170	52.5	68	4	US-10-663-450-60
20	117	36.1	203	4	US-10-437-963-151695
21	117	36.1	203	5	US-10-732-923-13694
22	117	36.1	203	5	US-10-732-923-13695
23	116	35.8	168	3	US-09-334-455-102
24	116	35.8	168	4	US-10-325-066A-206
25	116	35.8	168	4	US-10-374-780A-2356
26	116	35.8	168	5	US-10-732-923-13692
27	116	35.8	168	5	US-10-225-066A-206

28 111.5 34.4 192 3 US-09-772-656-2 Sequence 2, Appli
29 111.5 34.4 192 3 US-09-772-656-6 Sequence 6, Appli
30 111.5 34.4 192 3 US-09-772-656-10 Sequence 10, Appli
31 111.5 34.4 192 4 US-10-396-199A-2 Sequence 2, Appli
32 111.5 34.4 192 4 US-10-396-199A-6 Sequence 6, Appli
33 111.5 34.4 192 4 US-10-396-199A-10 Sequence 10, Appli
34 111.5 34.4 192 4 US-10-425-115-187302 Sequence 187302,
35 109 33.6 109 5 US-10-732-923-13697 Sequence 13697, A
36 109 33.6 196 4 US-10-425-115-197786 Sequence 197786,
37 108 33.3 176 4 US-10-437-963-164243 Sequence 164243,
38 106.5 32.9 170 4 US-10-310-154-565 Sequence 565, App
39 104.5 32.3 191 4 US-10-424-599-177263 Sequence 177263,
40 104 32.1 181 5 US-10-732-923-450 Sequence 450, App
41 101 31.2 217 4 US-10-425-115-197783 Sequence 197783,
42 100 30.9 143 5 US-10-856-499-1930 Sequence 1930, Ap
43 99 30.6 114 4 US-10-425-115-246430 Sequence 246430,
44 99 30.6 360 4 US-10-437-963-165206 Sequence 165206,
45 98 30.2 149 5 US-10-732-923-13693 Sequence 13693, A

ALIGNMENTS

RESULT 1

US-09-816-277-4
; Sequence 4, Application US/09816277
; Publication No. US20010034045A1
; GENERAL INFORMATION:
; APPLICANT: Penttila, Merja E.
; APPLICANT: Ward, Michael
; APPLICANT: Wang, Huaming
; APPLICANT: Valkonen, Mari J.
; APPLICANT: Saloheimo, Markku
; TITLE OF INVENTION: Increased Production of Secreted
; TITLE OF INVENTION: Proteins by Recombinant Eukaryotic Cells
; FILE REFERENCE: GC590-2
; CURRENT APPLICATION NUMBER: US/09/816,277
; CURRENT FILING DATE: 2001-03-23
; PRIOR APPLICATION NUMBER: US 09/534,692
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 349
; TYPE: PRT
; ORGANISM: Aspergillus nidulans
US-09-816-277-4

Query Match 100.0%; Score 324; DB 3; Length 349;
Best Local Similarity 100.0%; Pred. No. 5.1e-25;
Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KKPAAKRSWGQELPVKTNLPKRAKTEDEKQRIERVLNRAAAQTSRRKRLME 60
DB 53 KKPAAKRSWGQELPVKTNLPKRAKTEDEKQRIERVLNRAAAQTSRRKRLME 112
QY 61 KLES 64
DB 113 KLES 116

RESULT 2

US-09-816-277-6
; Sequence 6, Application US/09816277
; Publication No. US20010034045A1
; GENERAL INFORMATION:
; APPLICANT: Penttila, Merja E.
; APPLICANT: Ward, Michael
; APPLICANT: Wang, Huaming
; APPLICANT: Valkonen, Mari J.
; APPLICANT: Saloheimo, Markku
; TITLE OF INVENTION: Increased Production of Secreted
; TITLE OF INVENTION: Proteins by Recombinant Eukaryotic Cells

```
; FILE REFERENCE: GC590-2
; CURRENT APPLICATION NUMBER: US/09/816,277
; CURRENT FILING DATE: 2001-03-23
; PRIOR APPLICATION NUMBER: US 09/534,692
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 349
; TYPE: PRT
; ORGANISM: Aspergillus nidulans
US-09-816-277-6

Query Match      100.0%; Score 324; DB 3; Length 349;
Best Local Similarity 100.0%; Pred. No. 5.1e-25;
Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KKPAAKRRKSWGQELPVPKTNLPKRKRAKTEDEKEQRRRIERVLRNRAAAQTSRKRKLEME 60
   |||||
Db 53 KKPAAKRRKSWGQELPVPKTNLPKRKRAKTEDEKEQRRRIERVLRNRAAAQTSRKRKLEME 112
   |||||
QY 61 KLES 64
   |||||
Db 113 KLES 116

RESULT 3
US-10-663-450-4
; Sequence 4, Application US/10663450
; Publication No. US20040186070A1
; GENERAL INFORMATION:
; APPLICANT: Penttila, Merja E.
; APPLICANT: Ward, Michael
; APPLICANT: Wang, Huaming
; APPLICANT: Valkonen, Mari J.
; APPLICANT: Saloheimo, Markku
; TITLE OF INVENTION: Increased Production of Secreted
; CURRENT APPLICATION NUMBER: US/10/663,450
; CURRENT FILING DATE: 2003-09-15
; PRIOR APPLICATION NUMBER: US 09/534,692
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 349
; TYPE: PRT
; ORGANISM: Aspergillus nidulans
US-10-663-450-4

Query Match      100.0%; Score 324; DB 4; Length 349;
Best Local Similarity 100.0%; Pred. No. 5.1e-25;
Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KKPAAKRRKSWGQELPVPKTNLPKRKRAKTEDEKEQRRRIERVLRNRAAAQTSRKRKLEME 60
   |||||
Db 53 KKPAAKRRKSWGQELPVPKTNLPKRKRAKTEDEKEQRRRIERVLRNRAAAQTSRKRKLEME 112
   |||||
QY 61 KLES 64
   |||||
Db 113 KLES 116

RESULT 4
US-10-663-450-6
; Sequence 6, Application US/10663450
; Publication No. US20040186070A1
; GENERAL INFORMATION:
; APPLICANT: Penttila, Merja E.
; APPLICANT: Ward, Michael
; APPLICANT: Wang, Huaming
; APPLICANT: Valkonen, Mari J.
; APPLICANT: Saloheimo, Markku
; TITLE OF INVENTION: Increased Production of Secreted
; CURRENT APPLICATION NUMBER: US/10/663,450
; CURRENT FILING DATE: 2003-09-15
; PRIOR APPLICATION NUMBER: US 09/534,692
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 349
; TYPE: PRT
; ORGANISM: Aspergillus nidulans
US-10-663-450-6

Query Match      100.0%; Score 324; DB 3; Length 349;
Best Local Similarity 100.0%; Pred. No. 5.1e-25;
Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KKPAAKRRKSWGQELPVPKTNLPKRKRAKTEDEKEQRRRIERVLRNRAAAQTSRKRKLEME 60
   |||||
Db 53 KKPAAKRRKSWGQELPVPKTNLPKRKRAKTEDEKEQRRRIERVLRNRAAAQTSRKRKLEME 112
   |||||
QY 61 KLES 64
   |||||
Db 113 KLES 116

RESULT 5
US-09-816-277-19
; Sequence 19, Application US/09816277
; Publication No. US20010034045A1
; GENERAL INFORMATION:
; APPLICANT: Penttila, Merja E.
; APPLICANT: Ward, Michael
; APPLICANT: Wang, Huaming
; APPLICANT: Valkonen, Mari J.
; APPLICANT: Saloheimo, Markku
; TITLE OF INVENTION: Increased Production of Secreted
; TITLE OF INVENTION: Proteins by Recombinant Eukaryotic Cells
; FILE REFERENCE: GC590-2
; CURRENT APPLICATION NUMBER: US/09/816,277
; CURRENT FILING DATE: 2001-03-23
; PRIOR APPLICATION NUMBER: US 09/534,692
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 19
; LENGTH: 342
; TYPE: PRT
; ORGANISM: Aspergillus niger
US-09-816-277-19

Query Match      97.8%; Score 317; DB 3; Length 342;
Best Local Similarity 96.9%; Pred. No. 2.6e-24;
Matches 62; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 KKPAAKRRKSWGQELPVPKTNLPKRKRAKTEDEKEQRRRIERVLRNRAAAQTSRKRKLEME 60
   |||||
Db 47 KKPAAKRRKSWGQELPVPKTNLPKRKRAKTEDEKEQRRRIERVLRNRAAAQTSRKRKLEME 106
   |||||
QY 61 KLES 64
   |||||
Db 107 KLEN 110

RESULT 6
US-10-663-450-19
; Sequence 19, Application US/10663450
; Publication No. US20040186070A1
; GENERAL INFORMATION:
; APPLICANT: Penttila, Merja E.
```

Publication No.	US20040186070A1
GENERAL INFORMATION:	
APPLICANT: Penttila, Merja E.	
APPLICANT: Ward, Michael	
APPLICANT: Wang, Huaming	
APPLICANT: Valkonen, Markku	
APPLICANT: Saloheimo, Mari J.	
APPLICANT: Saloheimo, Markku	
TITLE OF INVENTION: Increased Production of Secreted	
PROTEINS BY RECOMBINANT EUKARYOTIC CELLS	
FILE REFERENCE: GC590-2	
CURRENT APPLICATION NUMBER: US/10/663,450	
CURRENT FILING DATE: 2003-09-15	
PRIOR APPLICATION NUMBER: US 09/534,692	
PRIOR FILING DATE: 2000-03-24	
NUMBER OF SEQ ID NOS: 63	
SOFTWARE: FastSeq for Windows Version 4.0	
SEQ ID NO 19	
LENGTH: 342	
TYPE: PRT	
ORGANISM: Aspergillus niger	
US-10-663-450-19	
Query Match	97.8%; Score 317; DB 4; Length 342;
Best Local Similarity	96.9%; Pred. No. 2.6e-24;
Matches	62; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY	1 KKPAAKRSWGQELPVPKTNLPKRKRAKTEDEKQRIERVLNRNRAAAQTSRKRKLEME 60
DB	47 KKPAAKRSWGQELPVPKTNLPKRKRAKTEDEKQRIERVLNRNRAAAQTSRKRKLEME 106
QY	61 KLES 64
DB	107 KLEN 110
RESULT 7	
US-09-816-277-16	
Sequence 16, Application US/09816277	
Publication No. US20010034045A1	
GENERAL INFORMATION:	
APPLICANT: Penttila, Merja E.	
APPLICANT: Ward, Michael	
APPLICANT: Wang, Huaming	
APPLICANT: Valkonen, Mari J.	
APPLICANT: Saloheimo, Markku	
TITLE OF INVENTION: Increased Production of Secreted	
PROTEINS BY RECOMBINANT EUKARYOTIC CELLS	
FILE REFERENCE: GC590-2	
CURRENT APPLICATION NUMBER: US/09/816,277	
CURRENT FILING DATE: 2001-03-23	
PRIOR APPLICATION NUMBER: US 09/534,692	
PRIOR FILING DATE: 2000-03-24	
NUMBER OF SEQ ID NOS: 63	
SOFTWARE: FastSeq for Windows Version 4.0	
SEQ ID NO 16	
LENGTH: 386	
TYPE: PRT	
ORGANISM: Aspergillus niger	
US-09-816-277-16	
Query Match	97.8%; Score 317; DB 3; Length 386;
Best Local Similarity	96.9%; Pred. No. 2.9e-24;
Matches	62; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY	1 KKPAAKRSWGQELPVPKTNLPKRKRAKTEDEKQRIERVLNRNRAAAQTSRKRKLEME 60
DB	91 KKPAAKRSWGQELPVPKTNLPKRKRAKTEDEKQRIERVLNRNRAAAQTSRKRKLEME 150
QY	61 KLES 64
DB	151 KLEN 154
RESULT 8	
US-10-663-450-16	
Sequence 16, Application US/10663450	
Publication No. US20040186070A1	
GENERAL INFORMATION:	
APPLICANT: Penttila, Merja E.	
APPLICANT: Ward, Michael	
APPLICANT: Wang, Huaming	
APPLICANT: Valkonen, Markku	
APPLICANT: Saloheimo, Mari J.	
APPLICANT: Saloheimo, Markku	
TITLE OF INVENTION: Increased Production of Secreted	
PROTEINS BY RECOMBINANT EUKARYOTIC CELLS	
FILE REFERENCE: GC590-2	
CURRENT APPLICATION NUMBER: US/10/663,450	
CURRENT FILING DATE: 2003-09-15	
PRIOR APPLICATION NUMBER: US 09/534,692	
PRIOR FILING DATE: 2000-03-24	
NUMBER OF SEQ ID NOS: 63	
SOFTWARE: FastSeq for Windows Version 4.0	
SEQ ID NO 19	
LENGTH: 342	
TYPE: PRT	
ORGANISM: Aspergillus niger	
US-10-663-450-19	
Query Match	97.8%; Score 317; DB 4; Length 342;
Best Local Similarity	96.9%; Pred. No. 2.6e-24;
Matches	62; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY	1 KKPAAKRSWGQELPVPKTNLPKRKRAKTEDEKQRIERVLNRNRAAAQTSRKRKLEME 60
DB	47 KKPAAKRSWGQELPVPKTNLPKRKRAKTEDEKQRIERVLNRNRAAAQTSRKRKLEME 106
QY	61 KLES 64
DB	107 KLEN 110
RESULT 9	
US-09-816-277-2	
Sequence 2, Application US/09816277	
Publication No. US20010034045A1	
GENERAL INFORMATION:	
APPLICANT: Penttila, Merja E.	
APPLICANT: Ward, Michael	
APPLICANT: Wang, Huaming	
APPLICANT: Valkonen, Mari J.	
APPLICANT: Saloheimo, Markku	
TITLE OF INVENTION: Increased Production of Secreted	
PROTEINS BY RECOMBINANT EUKARYOTIC CELLS	
FILE REFERENCE: GC590-2	
CURRENT APPLICATION NUMBER: US/09/816,277	
CURRENT FILING DATE: 2001-03-23	
PRIOR APPLICATION NUMBER: US 09/534,692	
PRIOR FILING DATE: 2000-03-24	
NUMBER OF SEQ ID NOS: 63	
SOFTWARE: FastSeq for Windows Version 4.0	
SEQ ID NO 2	
LENGTH: 451	
TYPE: PRT	
ORGANISM: Trichoderma reesei	
US-09-816-277-2	
Query Match	87.3%; Score 283; DB 3; Length 451;
Best Local Similarity	87.3%; Pred. No. 1e-20;
Matches	55; Conservative 3; Mismatches 5; Indels 0; Gaps 0;
QY	1 KKPAAKRSWGQELPVPKTNLPKRKRAKTEDEKQRIERVLNRNRAAAQTSRKRKLEME 60
DB	85 KKPAAKRSWGQELPVPKTNLPKRKRAKTEDEKQRIERVLNRNRAAAQTSRKRKLEME 144
QY	61 KLE 63
DB	145 ALK 147

```
RESULT 10
US-09-816-277-5
; Sequence 5, Application US/09816277
; Publication No. US20010034045A1
; GENERAL INFORMATION:
; APPLICANT: Penttila, Merja E.
; APPLICANT: Wang, Michael
; APPLICANT: Wang, Huaming
; APPLICANT: Valkonen, Mari J.
; APPLICANT: Saloheimo, Markku
; TITLE OF INVENTION: Increased Production of Secreted
; FILE REFERENCE: GC590-2
; CURRENT FILING DATE: 2003-03-23
; CURRENT APPLICATION NUMBER: US 09/534,692
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 451
; TYPE: PRT
; ORGANISM: Trichoderma reesei
US-09-816-277-5

Query Match      87.3%; Score 283; DB 3; Length 451;
Best Local Similarity 87.3%; Pred. No. 1e-20;
Matches 55; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 1 KKPVKRRKSWGQGLPVPKTNLPKRKAKTEDEKEQRRVERVLNRNRAAAQTSRERKLEME 60
Db 85 KKPVKRRKSWGQGLPVPKTNLPKRKAKTEDEKEQRRVERVLNRNRAAAQTSRERKLEVE 144
QY 61 KLE 63
Db 145 ALE 147

RESULT 11
US-10-663-450-2
; Sequence 2, Application US/10663450
; Publication No. US20040186070A1
; GENERAL INFORMATION:
; APPLICANT: Penttila, Merja E.
; APPLICANT: Wang, Michael
; APPLICANT: Wang, Huaming
; APPLICANT: Valkonen, Mari J.
; APPLICANT: Saloheimo, Markku
; TITLE OF INVENTION: Increased Production of Secreted
; FILE REFERENCE: GC590-2
; CURRENT FILING DATE: 2003-09-15
; CURRENT APPLICATION NUMBER: US 09/534,692
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 451
; TYPE: PRT
; ORGANISM: Trichoderma reesei
US-10-663-450-2

Query Match      87.3%; Score 283; DB 4; Length 451;
Best Local Similarity 87.3%; Pred. No. 1e-20;
Matches 55; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 1 KKPVKRRKSWGQGLPVPKTNLPKRKAKTEDEKEQRRVERVLNRNRAAAQTSRERKLEME 60
Db 85 KKPVKRRKSWGQGLPVPKTNLPKRKAKTEDEKEQRRVERVLNRNRAAAQTSRERKLEVE 144
QY 61 KLE 63
Db 145 ALE 147

RESULT 12
US-10-663-450-5
; Sequence 5, Application US/10663450
; Publication No. US20040186070A1
; GENERAL INFORMATION:
; APPLICANT: Penttila, Merja E.
; APPLICANT: Wang, Michael
; APPLICANT: Wang, Huaming
; APPLICANT: Valkonen, Mari J.
; APPLICANT: Saloheimo, Markku
; TITLE OF INVENTION: Increased Production of Secreted
; FILE REFERENCE: GC590-2
; CURRENT FILING DATE: 2003-09-15
; CURRENT APPLICATION NUMBER: US 09/534,692
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 451
; TYPE: PRT
; ORGANISM: Trichoderma reesei
US-10-663-450-5

Query Match      87.3%; Score 283; DB 4; Length 451;
Best Local Similarity 87.3%; Pred. No. 1e-20;
Matches 55; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 1 KKPVKRRKSWGQGLPVPKTNLPKRKAKTEDEKEQRRVERVLNRNRAAAQTSRERKLEME 60
Db 85 KKPVKRRKSWGQGLPVPKTNLPKRKAKTEDEKEQRRVERVLNRNRAAAQTSRERKLEVE 144
QY 61 KLE 63
Db 145 ALE 147

RESULT 13
US-09-533-029-104
; Sequence 104, Application US/09533029
; Publication No. US20030046723A1
; GENERAL INFORMATION:
; APPLICANT: Heard, Jacqueline
; APPLICANT: Broun, Pierre
; APPLICANT: Riechmann, Jose-Luis
; APPLICANT: Keddle, James
; APPLICANT: Pineda, Omaira
; APPLICANT: Adam, Luc
; APPLICANT: Samaha, Raymond
; APPLICANT: Zhang, James
; APPLICANT: Yu, Guo-Liang
; APPLICANT: Ratcliffe, Oliver
; APPLICANT: Pilgrim, Marsha
; APPLICANT: Jiang, Cai-Zhong
; APPLICANT: Reuber, Lynne
; TITLE OF INVENTION: DISEASE-INDUCED POLYNUCLEOTIDES
; FILE REFERENCE: MBI-010
; CURRENT FILING DATE: 2000-03-22
; CURRENT APPLICATION NUMBER: US/09/533,029
; EARLIER FILING DATE: 2000-03-22
; EARLIER APPLICATION NUMBER: 60/125,814
; EARLIER FILING DATE: 1999-03-23
; NUMBER OF SEQ ID NOS: 121
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 104
; LENGTH: 409
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; OTHER INFORMATION: G1034
```

RESULT 15
US-10-412-699B-684
; Sequence 684, Application US/10412699B
; Publication No. US20040045049A1

Search completed: November 22, 2005, 15:54:37
Job time : 113 secs

THIS PAGE BLANK (USPTO)

Result No.	Score	Query Match	Length	DB	ID	Description	
1	77	23.8	376	1	US-10-971-994-2	Sequence 2, Appli	
2	66.5	20.6	605	1	US-10-131-826A-160	Sequence 160, Appl	
3	65	20.1	431	7	US-11-169-013-2	Sequence 2, Appli	
4	62.5	19.3	454	7	US-11-169-013-1	Sequence 1, Appli	
5	59	18.3	747	7	US-11-182-592-2	Sequence 2, Appli	
6	58.5	18.1	903	1	US-10-689-742-142	Sequence 142, Appl	
7	57	17.6	312	7	US-10-986-501-124	Sequence 124, Appl	
8	57	17.6	400	7	US-11-018-018-7	Sequence 7, Appli	
9	56.5	17.5	180	1	US-10-689-742-84	Sequence 84, Appl	
10	56.5	17.5	1618	1	US-10-384-645-2	Sequence 2, Appli	
11	54	16.7	21	1	US-10-995-951A-33	Sequence 33, Appli	
12	54	16.7	170	1	US-10-689-742-16	Sequence 16, Appl	
13	53.5	16.6	569	1	US-10-632-150-2	Sequence 2, Appli	
14	51.5	15.9	615	1	US-10-582-545-14	Sequence 14, Appl	
15	51.5	15.9	879	7	US-11-022-562-340	Sequence 340, Appl	
16	51	15.8	2897	1	US-10-499-715-2	Sequence 2, Appli	
17	50.5	15.6	296	7	US-11-078-280-2	Sequence 2, Appli	
18	50.5	15.6	327	7	US-11-090-878-16	Sequence 16, Appl	
19	50.5	15.6	684	7	US-11-012-762-8	Sequence 8, Appli	
20	50	15.5	790	1	US-10-131-826A-204	Sequence 204, Appl	
21	49.5	15.3	128	7	US-11-105-864-4	Sequence 4, Appli	
22	49	15.2	188	1	US-10-131-826A-2	Sequence 2, Appli	
23	49	15.2	194	7	US-11-074-176-82	Sequence 82, Appl	
24	49	15.2	625	1	US-10-510-386-6	Sequence 6, Appli	
25	49	15.2	1302	7	US-11-004-057-6	Sequence 6, Appli	

```

; APPLICANT: DeForge,Laura
; APPLICANT: Desnoyers,Luc
; APPLICANT: Filvaroff,Ellen
; APPLICANT: Gao,Wei-Qiang
; APPLICANT: Gerritsen,Mary E.
; APPLICANT: Goddard,Audrey
; APPLICANT: Godowski,Paul J.
; APPLICANT: Gurney,Austin L.
; APPLICANT: Sherwood,Steven
; APPLICANT: Smith,Victoria
; APPLICANT: Stewart,Timothy A.
; APPLICANT: Tumas,Daniel
; APPLICANT: Watanabe,Colin K
; APPLICANT: Wood,William
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: SECRETED AND TRANSMEMBRANE POLYPEPTIDES AND NUCLEIC
; FILE REFERENCE: P3330RIC128
; CURRENT APPLICATION NUMBER: US/10/131,826A
; CURRENT FILING DATE: 2002-04-24
; PRIOR APPLICATION NUMBER: 60/049911
; PRIOR FILING DATE: 1997-06-18
; PRIOR APPLICATION NUMBER: 60/056974
; PRIOR FILING DATE: 1997-08-26
; PRIOR APPLICATION NUMBER: 60/059113
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059115
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059117
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059122
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059184
; PRIOR FILING DATE: 1997-09-17
; PRIOR APPLICATION NUMBER: 60/059263
; PRIOR FILING DATE: 1997-09-18
; PRIOR APPLICATION NUMBER: 60/059352
; PRIOR FILING DATE: 1997-09-19
; PRIOR APPLICATION NUMBER: 60/059588
; PRIOR FILING DATE: 1997-09-19
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 550
; SEQ ID NO 160
; LENGTH: 605
; TYPE: PRT
; ORGANISM: Homo Sapien
US-10-131-826A-160

Query Match 20.6%; Score 66.5; DB 1; Length 605;
Best Local Similarity 34.1%; Pred.No.0.41;
Matches 15; Conservative 9; Mismatches 15; Indels 5; Gaps

QY 15 LPPEKTNLPKPKAKTEDEKQRRVRLRNRAAQSSRRKRL 58
| | | | | : | | | | : | | | | : | | | |
Db 151 LVPPWHMPPREEEEEEEEKEKEVEK-----QEEEEEL 189

RESULT 3
US-11-169-013-2
; Sequence 2, Application US/11169013
; Publication No. US20050244971A1
; GENERAL INFORMATION:
; APPLICANT: Korea Kumho Petrochemical Co., Ltd.
; TITLE OF INVENTION: TRANSGENIC PLANTS WITH ENHANCED STRESS TOLERANCE
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/11/169,013
; CURRENT FILING DATE: 2005-06-29
; NUMBER OF SEQ ID NOS: 34
; SOFTWARE: Kopatentin 1.71
; SEQ ID NO 2
; LENGTH: 431
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana

```



```

Db          701 KEELARL 707

RESULT 6
US-10-689-742-142
; Sequence 142, Application US/10689742
; Publication No. US20050250180A1
; GENERAL INFORMATION:
; APPLICANT: Jacobs, Kenneth
; APPLICANT: McCoy, John M
; APPLICANT: LaVallie, Edward R
; APPLICANT: Racie, Lisa A
; APPLICANT: Evans, Cheryl
; APPLICANT: Merberg, David
; APPLICANT: Treacy, Maurice
; APPLICANT: Spaulding, Vikki
; TITLE OF INVENTION: SCRPTED PROTEINS AND POLYNUCLEOTIDES ENCODING THEM
; FILE REFERENCE: 00766.000091.10
; CURRENT APPLICATION NUMBER: US/10/689,742
; CURRENT FILING DATE: 2003-10-22
; PRIOR APPLICATION NUMBER: 09/746,783
; PRIOR FILING DATE: 2000-12-21
; NUMBER OF SEQ ID NOS: 231
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 142
; LENGTH: 903
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-689-742-142

Query Match      18.1%; Score 58.5; DB 1; Length 903;
Best Local Similarity 30.5%; Pred. No. 4.9;
Matches 18; Conservative 8; Mismatches 26; Indels 7; Gaps 1;

QY      6 KKRSGQVLPEPTNLPPEKRAKTEDEKEORRVERVLNRPRAQSSPERKRLSEVALE 64
Db      51 KKR-----PEPVP IQGNNNGRTTDLKQOSTRESWVSPKRGSLSSKDNIRQAIE 102

RESULT 7
US-10-986-501-124
; Sequence 124, Application US/10986501
; Publication No. US20050244845A1
; GENERAL INFORMATION:
; APPLICANT: Ruben et al.
; TITLE OF INVENTION: 90 Human Secreted Proteins
; FILE REFERENCE: PZ013P2C1
; CURRENT APPLICATION NUMBER: US/10/986,501
; CURRENT FILING DATE: 2004-11-12
; PRIOR APPLICATION NUMBER: US/10/621,363
; PRIOR FILING DATE: 2003-07-18
; PRIOR APPLICATION NUMBER: 09/969,730
; PRIOR FILING DATE: 2001-10-06
; PRIOR APPLICATION NUMBER: 09/774,639
; PRIOR FILING DATE: 2001-02-01
; PRIOR APPLICATION NUMBER: 60/238,291
; PRIOR FILING DATE: 2000-10-06
; PRIOR APPLICATION NUMBER: 09/244,112
; PRIOR FILING DATE: 1999-02-04
; PRIOR APPLICATION NUMBER: PCT/US98/16235
; PRIOR FILING DATE: 1998-08-04
; PRIOR APPLICATION NUMBER: 60/056,371
; PRIOR FILING DATE: 1997-08-19
; PRIOR APPLICATION NUMBER: 60/056,732
; PRIOR FILING DATE: 1997-08-19
; PRIOR APPLICATION NUMBER: 60/056,366
; PRIOR FILING DATE: 1997-08-19
; PRIOR APPLICATION NUMBER: 60/056,364
; PRIOR FILING DATE: 1997-08-19
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 373
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 124

```



```
; NUMBER OF SEQ ID NOS: 90
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 569
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-632-150-2

Query Match      16.6%; Score 53.5; DB 1; Length 569;
Best Local Similarity 28.8%; Pred. No. 10;
Matches 15; Conservative 8; Mismatches 26; Indels 3; Gaps 1;

QY 7 KKKSQGVLP---PEPKTNLPKRKRAKTEDEKEQRRVERVLRNRAAQSSRR 55
Db 208 ERKGQYLFKNKPPDGNAPPNSFYRALYPKIIDTETIESNWRGRSLQR 259

RESULT 14
US-10-982-545-14
; Sequence 14, Application US/10982545
; Publication No. US20050244890A1
; GENERAL INFORMATION:
; APPLICANT: Davies, Huw Alun
; APPLICANT: McGuire, James
; APPLICANT: Simonsen, Anja Hviid
; APPLICANT: Blennow, Kaj
; APPLICANT: Podust, Vladimir
; APPLICANT: Ciphergen Biosystems, Inc.
; TITLE OF INVENTION: Biomarkers for Alzheimer's Disease
; FILE REFERENCE: 016866-011550US
; CURRENT APPLICATION NUMBER: US/10/982,545
; PRIOR FILING DATE: 2004-11-06
; PRIOR APPLICATION NUMBER: US 60/518,360
; PRIOR FILING DATE: 2003-11-07
; PRIOR APPLICATION NUMBER: US 60/526,753
; PRIOR FILING DATE: 2003-12-02
; PRIOR APPLICATION NUMBER: US 60/546,423
; PRIOR FILING DATE: 2004-02-19
; PRIOR APPLICATION NUMBER: US 60/547,250
; PRIOR FILING DATE: 2004-02-23
; PRIOR APPLICATION NUMBER: US 60/558,896
; PRIOR FILING DATE: 2004-04-02
; PRIOR APPLICATION NUMBER: US 60/572,617
; PRIOR FILING DATE: 2004-05-18
; PRIOR APPLICATION NUMBER: US 60/586,503
; PRIOR FILING DATE: 2004-07-08
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 14
; LENGTH: 615
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; OTHER INFORMATION: nerve growth factor inducible Neurosecretory
; OTHER INFORMATION: protein vaccinia virus growth factor (VGF)
; OTHER INFORMATION: precursor
; FEATURE:
; NAME/KEY: SIGNAL
; LOCATION: (1)..(22)
; OTHER INFORMATION: signal peptide
; FEATURE:
; NAME/KEY: PEPTIDE
; LOCATION: (23)..(615)
; OTHER INFORMATION: nerve growth factor inducible Neurosecretory
; OTHER INFORMATION: protein vaccinia virus growth factor (VGF)
; FEATURE:
; NAME/KEY: PEPTIDE
; LOCATION: (373)..(417)
; OTHER INFORMATION: biomarker peptide 4808 Da, 4812 Da (Q10),
; OTHER INFORMATION: processed fragment of nerve growth factor
; OTHER INFORMATION: inducible Neurosecretory protein vaccinia virus
; OTHER INFORMATION: growth factor (VGF)
US-10-982-545-14
```

```
Query Match      15.9%; Score 51.5; DB 1; Length 615;
Best Local Similarity 31.5%; Pred. No. 19;
Matches 17; Conservative 8; Mismatches 14; Indels 15; Gaps 3;

QY 7 KKKSQGVLPPEPKTNLPKR---KRAKTEDEKEQRR-----VERVLRNR 46
Db 562 RRRHYHALP-PSRRYGRQAARQAEEAEARRLQEQEELENYIEHVLR 614
```

```
RESULT 15
US-11-022-562-340
; Sequence 340, Application US/11022562
; Publication No. US20050249742A1
; GENERAL INFORMATION:
; APPLICANT: Rupprecht, Ruth M.
; APPLICANT: Shisong, Jiang
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR MODULATING
; TITLE OF INVENTION: A CYTOTOXIC T LYMPHOCYTE IMMUNE RESPONSE
; FILE REFERENCE: DFN-043CN
; CURRENT APPLICATION NUMBER: US/11/022,562
; CURRENT FILING DATE: 2004-12-22
; PRIOR APPLICATION NUMBER: PCT/US03/20322
; PRIOR FILING DATE: 2003-06-27
; PRIOR APPLICATION NUMBER: 60/392718
; PRIOR FILING DATE: 2002-06-27
; NUMBER OF SEQ ID NOS: 340
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 340
; LENGTH: 879
; TYPE: PRT
; ORGANISM: Simian Immunodeficiency Virus
US-11-022-562-340
```

```
Query Match      15.9%; Score 51.5; DB 7; Length 879;
Best Local Similarity 28.8%; Pred. No. 28;
Matches 19; Conservative 12; Mismatches 22; Indels 13; Gaps 3;

QY 10 SWG-----QVLPEPKTNLPKRKRAKTEDEKEQRRVERVLRNRRRA----AQSSRRERKR 57
Db 608 AWGCAFRQVCHTTPWPNTSLTPKWDNETQWEB-RKVDFFLEENITALPEEAQIQQEKNM 666

QY 58 LEVEAL 63
Db 667 YELQKL 672
```

Search completed: November 22, 2005, 15:54:45
Job time : 5 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 22, 2005, 15:38:17 ; Search time 111 Seconds
(without alignments)
240.911 Million cell updates/sec

Title: US-10-663-450-5_COPY_84_147

Perfect score: 323

Sequence: 1 EKKPVKRSWGQVLPEPT.....NRAAQSRRKRLEVALE 64

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Databae : 1: /cgn2_6/prodata/1/pubpaa/US07_PUBCOMB.pap:*
2: /cgn2_6/prodata/1/pubpaa/US08_PUBCOMB.pap:*
3: /cgn2_6/prodata/1/pubpaa/US09_PUBCOMB.pap:*
4: /cgn2_6/prodata/1/pubpaa/US10A_PUBCOMB.pap:*
5: /cgn2_6/prodata/1/pubpaa/US10B_PUBCOMB.pap:*
6: /cgn2_6/prodata/1/pubpaa/US11_PUBCOMB.pap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	323	100.0	451	3	US-09-816-277-2
2	323	100.0	451	3	US-09-816-277-5
3	323	100.0	451	4	US-10-663-450-2
4	323	100.0	451	4	US-10-663-450-5
5	307	95.0	409	3	US-09-533-029-104
6	307	95.0	409	4	US-10-295-403-158
7	307	95.0	409	4	US-10-412-699B-684
8	292	90.4	342	3	US-09-816-277-19
9	292	90.4	342	4	US-10-663-450-19
10	292	90.4	386	3	US-09-816-277-16
11	292	90.4	386	4	US-10-663-450-16
12	288	89.2	349	3	US-09-816-277-4
13	288	89.2	349	3	US-09-816-277-6
14	288	89.2	349	4	US-10-663-450-4
15	288	89.2	349	4	US-10-663-450-6
16	268	83.0	174	4	US-10-767-701-61238
17	175	54.2	68	3	US-09-816-277-60
18	175	54.2	68	4	US-10-663-450-60
19	175	54.2	200	4	US-10-369-493-21867
20	108	33.4	147	4	US-10-425-115-187302
21	108	33.4	192	3	US-09-772-656-2
22	108	33.4	192	3	US-09-772-656-6
23	108	33.4	192	3	US-09-772-656-10
24	108	33.4	192	4	US-10-396-199A-2
25	108	33.4	192	4	US-10-396-199A-6
26	108	33.4	192	4	US-10-396-199A-10
27	107	33.1	203	4	US-10-437-963-151695

28 107 33.1 203 5 US-10-732-923-13694 Sequence 13694, A
29 107 33.1 203 5 US-10-732-923-13695 Sequence 13695, A
30 106 32.8 168 3 US-09-934-453-102 Sequence 102, App
31 106 32.8 168 4 US-10-225-066A-206 Sequence 206, App
32 106 32.8 168 4 US-10-374-780A-2356 Sequence 2356, App
33 106 32.8 168 5 US-10-732-923-13692 Sequence 13692, A
34 106 32.8 168 5 US-10-225-066A-206 Sequence 206, App
35 103.5 32.0 191 4 US-10-424-599-177263 Sequence 177263, App
36 103 31.9 170 4 US-10-310-154-565 Sequence 565, App
37 102 31.6 109 5 US-10-732-923-13697 Sequence 13697, A
38 102 31.6 196 4 US-10-425-115-197786 Sequence 197786, App
39 101 31.3 665 4 US-10-425-115-190929 Sequence 190929, A
40 101 31.3 672 4 US-10-425-114-58469 Sequence 58469, A
41 99 30.7 143 5 US-10-856-499-1930 Sequence 1930, App
42 98 30.3 176 4 US-10-437-963-164243 Sequence 164243, App
43 97 30.0 181 5 US-10-732-923-450 Sequence 450, App
44 95 29.4 217 4 US-10-425-115-197783 Sequence 197783, App
45 94 29.1 114 4 US-10-425-115-246430 Sequence 246430, App

ALIGNMENTS

RESULT 1

US-09-816-277-2
; Sequence 2, Application US/09816277
; Publication No. US20010034045A1
; GENERAL INFORMATION:
; APPLICANT: Penttila, Merja E.
; APPLICANT: Ward, Michael
; APPLICANT: Wang, Huaming
; APPLICANT: Valkonen, Mari J.
; APPLICANT: Saloheimo, Markku
; TITLE OF INVENTION: Increased Production of Secreted
; TITLE OF INVENTION: Proteins by Recombinant Eukaryotic Cells
; FILE REFERENCE: GC590-2
; CURRENT APPLICATION NUMBER: US/09/816,277
; CURRENT FILING DATE: 2001-03-23
; PRIOR APPLICATION NUMBER: US 09/534,692
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 451
; TYPE: PRT
; ORGANISM: Trichoderma reesei
US-09-816-277-2

Query Match 100.0%; Score 323; DB 3; Length 451;
Best Local Similarity 100.0%; Pred. No. 8.9e-24;
Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EKKPVKRSWGQVLPEPTNLPPPKRAKTEDEKEORRVERVLNRRAAQSRRKRLEV 60
DB 84 EKKPVKRSWGQVLPEPTNLPPPKRAKTEDEKEORRVERVLNRRAAQSRRKRLEV 143
QY 61 EALE 64
DB 144 EALE 147

RESULT 2

US-09-816-277-5
; Sequence 5, Application US/09816277
; Publication No. US20010034045A1
; GENERAL INFORMATION:
; APPLICANT: Penttila, Merja E.
; APPLICANT: Ward, Michael
; APPLICANT: Wang, Huaming
; APPLICANT: Valkonen, Mari J.
; APPLICANT: Saloheimo, Markku
; TITLE OF INVENTION: Increased Production of Secreted
; TITLE OF INVENTION: Proteins by Recombinant Eukaryotic Cells

```
; FILE REFERENCE: GC590-2
; CURRENT APPLICATION NUMBER: US/09/816,277
; PRIOR FILING DATE: 2001-03-23
; PRIOR APPLICATION NUMBER: US 09/534,692
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 451
; TYPE: PRT
; ORGANISM: Trichoderma reesei
US-09-816-277-5

Query Match      100.0%; Score 323; DB 3; Length 451;
Best Local Similarity 100.0%; Pred. No. 8.9e-24;
Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EKKPVKKRKSQGVLPEPKTNLPPrKRKTEDEKEQRRVERVLNRRAAQSRRKRLEV 60
Db 84 EKKPVKKRKSQGVLPEPKTNLPPrKRKTEDEKEQRRVERVLNRRAAQSRRKRLEV 143

QY 61 EALE 64
Db 144 EALE 147

RESULT 3
US-10-663-450-2
; Sequence 2, Application US/10663450
; Publication No. US20040186070A1
; GENERAL INFORMATION:
; APPLICANT: Penttila, Merja E.
; APPLICANT: Ward, Michael
; APPLICANT: Wang, Huaming
; APPLICANT: Valkonen, Mari J.
; APPLICANT: Saloheimo, Markku
; TITLE OF INVENTION: Increased Production of Secreted
; FILE REFERENCE: GC590-2
; CURRENT APPLICATION NUMBER: US/10/663,450
; CURRENT FILING DATE: 2003-09-15
; PRIOR APPLICATION NUMBER: US 09/534,692
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 451
; TYPE: PRT
; ORGANISM: Trichoderma reesei
US-10-663-450-2

Query Match      100.0%; Score 323; DB 4; Length 451;
Best Local Similarity 100.0%; Pred. No. 8.9e-24;
Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EKKPVKKRKSQGVLPEPKTNLPPrKRKTEDEKEQRRVERVLNRRAAQSRRKRLEV 60
Db 84 EKKPVKKRKSQGVLPEPKTNLPPrKRKTEDEKEQRRVERVLNRRAAQSRRKRLEV 143

QY 61 EALE 64
Db 144 EALE 147

RESULT 4
US-10-663-450-5
; Sequence 5, Application US/10663450
; Publication No. US20040186070A1
; GENERAL INFORMATION:
; APPLICANT: Penttila, Merja E.
; APPLICANT: Ward, Michael
; APPLICANT: Wang, Huaming
; APPLICANT: Valkonen, Mari J.
```

```
; APPLICANT: Saloheimo, Markku
; TITLE OF INVENTION: Increased Production of Secreted
; FILE REFERENCE: GC590-2
; CURRENT APPLICATION NUMBER: US/10/663,450
; CURRENT FILING DATE: 2003-09-15
; PRIOR APPLICATION NUMBER: US 09/534,692
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 451
; TYPE: PRT
; ORGANISM: Trichoderma reesei
US-10-663-450-5

Query Match      100.0%; Score 323; DB 4; Length 451;
Best Local Similarity 100.0%; Pred. No. 8.9e-24;
Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EKKPVKKRKSQGVLPEPKTNLPPrKRKTEDEKEQRRVERVLNRRAAQSRRKRLEV 60
Db 84 EKKPVKKRKSQGVLPEPKTNLPPrKRKTEDEKEQRRVERVLNRRAAQSRRKRLEV 143

QY 61 EALE 64
Db 144 EALE 147

RESULT 5
US-09-533-029-104
; Sequence 104, Application US/09533029
; Publication No. US20030046723A1
; GENERAL INFORMATION:
; APPLICANT: Heard, Jacqueline
; APPLICANT: Broun, Pierre
; APPLICANT: Riechmann, Jose-Luis
; APPLICANT: Keddie, James
; APPLICANT: Pineda, Omaira
; APPLICANT: Adam, Luc
; APPLICANT: Samaha, Raymond
; APPLICANT: Zhang, James
; APPLICANT: Yu, Guo-Liang
; APPLICANT: Ratcliffe, Oliver
; APPLICANT: Pilgrim, Marsha
; APPLICANT: Jiang, Cai-Zhong
; APPLICANT: Reuber, Lynne
; TITLE OF INVENTION: DISEASE-INDUCED POLYNUCLEOTIDES
; FILE REFERENCE: MBI-010
; CURRENT APPLICATION NUMBER: US/09/533,029
; CURRENT FILING DATE: 2000-03-22
; EARLIER APPLICATION NUMBER: 60/125,814
; EARLIER FILING DATE: 1999-03-23
; NUMBER OF SEQ ID NOS: 121
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 104
; LENGTH: 409
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; OTHER INFORMATION: G1034
US-09-533-029-104

Query Match      95.0%; Score 307; DB 3; Length 409;
Best Local Similarity 93.8%; Pred. No. 3.1e-22;
Matches 60; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 EKKPVKKRKSQGVLPEPKTNLPPrKRKTEDEKEQRRVERVLNRRAAQSRRKRLEV 60
Db 66 EKKPVKKRKSQGVLPEPKTNLPPrKRKTEDEKEQRRVERVLNRRAAQSRRKRLEV 125

QY 61 EALE 64
Db 144 EALE 147
```



```
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 19
; LENGTH: 342
; TYPE: PRT
; ORGANISM: Aspergillus niger
US-09-816-277-19

Query Match
Best Local Similarity 90.4%; Score 292; DB 3; Length 342;
Matches 57; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 EKKPVKKRKSQGVLPPEPTNLPPRKRAKTEDEKEQRRVRLNRRAAQSRRKRLEV 60
    |||||
Db 46 EKKPVKKRKSQGVLPPEPTNLPPRKRAKTEDEKEQRRVRLNRRAAQSRRKRLEM 105
    |||||

QY 61 EALE 64
    |||
Db 106 EKLE 109

RESULT 9
US-10-663-450-19
; Sequence 19, Application US/10663450
; Publication No. US20040186070A1
; GENERAL INFORMATION:
; APPLICANT: Penttila, Merja E.
; APPLICANT: Ward, Michael
; APPLICANT: Wang, Huaming
; APPLICANT: Valkonen, Mari J.
; APPLICANT: Saloheimo, Markku
; TITLE OF INVENTION: Increased Production of Secreted
; FILE REFERENCE: GC590-2
; CURRENT APPLICATION NUMBER: US/10/663,450
; CURRENT FILING DATE: 2003-09-15
; PRIOR APPLICATION NUMBER: US 09/534,692
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 19
; LENGTH: 342
; TYPE: PRT
; ORGANISM: Aspergillus niger
US-10-663-450-19

Query Match
Best Local Similarity 90.4%; Score 292; DB 4; Length 342;
Matches 57; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 EKKPVKKRKSQGVLPPEPTNLPPRKRAKTEDEKEQRRVRLNRRAAQSRRKRLEV 60
    |||||
Db 46 EKKPVKKRKSQGVLPPEPTNLPPRKRAKTEDEKEQRRVRLNRRAAQSRRKRLEM 105
    |||||

QY 61 EALE 64
    |||
Db 106 EKLE 109

RESULT 10
US-09-816-277-16
; Sequence 16, Application US/09816277
; Publication No. US20010034045A1
; GENERAL INFORMATION:
; APPLICANT: Penttila, Merja E.
; APPLICANT: Ward, Michael
; APPLICANT: Wang, Huaming
; APPLICANT: Valkonen, Mari J.
; APPLICANT: Saloheimo, Markku
; TITLE OF INVENTION: Increased Production of Secreted
; FILE REFERENCE: GC590-2
; CURRENT APPLICATION NUMBER: US/09/816,277
```

```
; CURRENT FILING DATE: 2001-03-23
; PRIOR APPLICATION NUMBER: US 09/534,692
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16
; LENGTH: 386
; TYPE: PRT
; ORGANISM: Aspergillus niger
US-09-816-277-16

Query Match
Best Local Similarity 90.4%; Score 292; DB 3; Length 386;
Matches 57; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 EKKPVKKRKSQGVLPPEPTNLPPRKRAKTEDEKEQRRVRLNRRAAQSRRKRLEV 60
    |||||
Db 90 EKKPVKKRKSQGVLPPEPTNLPPRKRAKTEDEKEQRRVRLNRRAAQSRRKRLEM 149
    |||||

QY 61 EALE 64
    |||
Db 150 EKLE 153

RESULT 11
US-10-663-450-16
; Sequence 16, Application US/10663450
; Publication No. US20040186070A1
; GENERAL INFORMATION:
; APPLICANT: Penttila, Merja E.
; APPLICANT: Ward, Michael
; APPLICANT: Wang, Huaming
; APPLICANT: Valkonen, Mari J.
; APPLICANT: Saloheimo, Markku
; TITLE OF INVENTION: Increased Production of Secreted
; FILE REFERENCE: GC590-2
; CURRENT APPLICATION NUMBER: US/10/663,450
; CURRENT FILING DATE: 2003-09-15
; PRIOR APPLICATION NUMBER: US 09/534,692
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16
; LENGTH: 386
; TYPE: PRT
; ORGANISM: Aspergillus niger
US-10-663-450-16

Query Match
Best Local Similarity 90.4%; Score 292; DB 4; Length 386;
Matches 57; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 EKKPVKKRKSQGVLPPEPTNLPPRKRAKTEDEKEQRRVRLNRRAAQSRRKRLEV 60
    |||||
Db 90 EKKPVKKRKSQGVLPPEPTNLPPRKRAKTEDEKEQRRVRLNRRAAQSRRKRLEM 149
    |||||

QY 61 EALE 64
    |||
Db 150 EKLE 153

RESULT 12
US-09-816-277-4
; Sequence 4, Application US/09816277
; Publication No. US20010034045A1
; GENERAL INFORMATION:
; APPLICANT: Penttila, Merja E.
; APPLICANT: Ward, Michael
; APPLICANT: Wang, Huaming
; APPLICANT: Valkonen, Mari J.
; APPLICANT: Saloheimo, Markku
; TITLE OF INVENTION: Increased Production of Secreted
```


; TITLE OF INVENTION: Proteins by Recombinant Eukaryotic Cells
; FILE REFERENCE: GC590-2
; CURRENT APPLICATION NUMBER: US/09/816,277
; CURRENT FILING DATE: 2001-03-23
; PRIOR APPLICATION NUMBER: US 09/534,692
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 349
; TYPE: PRT
; ORGANISM: Aspergillus nidulans
US-09-816-277-4

Query Match 89.2%; Score 288; DB 3; Length 349;
Best Local Similarity 87.5%; Pred. No. 2e-20;
Matches 56; Conservative 3; Mismatches 5; Indels 0; Gaps 0;
QY 1 EKKPVKKRKSQGVLPPEPTNLPPPKRAKTEDEKEQRRVRLNRRAAQSSRRKRLEV 60
DB 52 EKKPAKRRKSWGQELPVPKTNLPKRAKTEDEKEQRRVRLNRRAAAQTSRRKRLEM 111
QY 61 EALE 64
DB 112 EKLE 115

RESULT 13
US-09-816-277-6
; Sequence 6, Application US/09816277
; Publication No. US20010034045A1
; GENERAL INFORMATION:
; APPLICANT: Penttila, Merja E.
; APPLICANT: Ward, Michael
; APPLICANT: Wang, Huaming
; APPLICANT: Valkonen, Mari J.
; APPLICANT: Saloheimo, Markku
; TITLE OF INVENTION: Increased Production of Secreted
; TITLE OF INVENTION: Proteins by Recombinant Eukaryotic Cells
; FILE REFERENCE: GC590-2
; CURRENT APPLICATION NUMBER: US/09/816,277
; CURRENT FILING DATE: 2001-03-23
; PRIOR APPLICATION NUMBER: US 09/534,692
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 349
; TYPE: PRT
; ORGANISM: Aspergillus nidulans
US-09-816-277-6

Query Match 89.2%; Score 288; DB 3; Length 349;
Best Local Similarity 87.5%; Pred. No. 2e-20;
Matches 56; Conservative 3; Mismatches 5; Indels 0; Gaps 0;
QY 1 EKKPVKKRKSQGVLPPEPTNLPPPKRAKTEDEKEQRRVRLNRRAAQSSRRKRLEV 60
DB 52 EKKPAKRRKSWGQELPVPKTNLPKRAKTEDEKEQRRVRLNRRAAAQTSRRKRLEM 111
QY 61 EALE 64
DB 112 EKLE 115

RESULT 14
US-10-663-450-4
; Sequence 4, Application US/10663450
; Publication No. US20040186070A1
; GENERAL INFORMATION:
; APPLICANT: Penttila, Merja E.
; APPLICANT: Ward, Michael
; APPLICANT: Wang, Huaming

; APPLICANT: Valkonen, Mari J.
; APPLICANT: Saloheimo, Markku
; TITLE OF INVENTION: Increased Production of Secreted
; TITLE OF INVENTION: Proteins by Recombinant Eukaryotic Cells
; FILE REFERENCE: GC590-2
; CURRENT APPLICATION NUMBER: US/10/663,450
; CURRENT FILING DATE: 2003-09-15
; PRIOR APPLICATION NUMBER: US 09/534,692
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 349
; TYPE: PRT
; ORGANISM: Aspergillus nidulans
US-10-663-450-4

Query Match 89.2%; Score 288; DB 4; Length 349;
Best Local Similarity 87.5%; Pred. No. 2e-20;
Matches 56; Conservative 3; Mismatches 5; Indels 0; Gaps 0;
QY 1 EKKPVKKRKSQGVLPPEPTNLPPPKRAKTEDEKEQRRVRLNRRAAQSSRRKRLEV 60
DB 52 EKKPAKRRKSWGQELPVPKTNLPKRAKTEDEKEQRRVRLNRRAAAQTSRRKRLEM 111
QY 61 EALE 64
DB 112 EKLE 115

RESULT 15
US-10-663-450-6
; Sequence 6, Application US/10663450
; Publication No. US20040186070A1
; GENERAL INFORMATION:
; APPLICANT: Penttila, Merja E.
; APPLICANT: Ward, Michael
; APPLICANT: Wang, Huaming
; APPLICANT: Valkonen, Mari J.
; APPLICANT: Saloheimo, Markku
; TITLE OF INVENTION: Increased Production of Secreted
; TITLE OF INVENTION: Proteins by Recombinant Eukaryotic Cells
; FILE REFERENCE: GC590-2
; CURRENT APPLICATION NUMBER: US/10/663,450
; CURRENT FILING DATE: 2003-09-15
; PRIOR APPLICATION NUMBER: US 09/534,692
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 349
; TYPE: PRT
; ORGANISM: Aspergillus nidulans
US-10-663-450-6

Query Match 89.2%; Score 288; DB 4; Length 349;
Best Local Similarity 87.5%; Pred. No. 2e-20;
Matches 56; Conservative 3; Mismatches 5; Indels 0; Gaps 0;
QY 1 EKKPVKKRKSQGVLPPEPTNLPPPKRAKTEDEKEQRRVRLNRRAAQSSRRKRLEV 60
DB 52 EKKPAKRRKSWGQELPVPKTNLPKRAKTEDEKEQRRVRLNRRAAAQTSRRKRLEM 111
QY 61 EALE 64
DB 112 EKLE 115

Search completed: November 22, 2005, 15:54:35
Job time : 117 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 22, 2005, 15:33:00 ; Search time 147.5 Seconds
(without alignments)
306.128 Million cell updates/sec

Title: US-10-663-450-5_COPY_84_147
Perfect score: 323
Sequence: 1 EKKPVKRSWGQVLPEPKT.....NRRAAQSRKRKRLEVALE 64

Scoring table:
BLOSUM62
Gapop 10.0 , Gapept 0.5

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Uniprot 05.80.*
1: uniprot_sprot.*
2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	323	100.0	451	2	Q8TFP3	TRIRE
2	319	98.8	430	2	Q7SHF0	NEUCR
3	296	91.6	556	2	Q51KW8	MAGGR
4	292	90.4	342	2	Q6W8X2	ASPNB
5	288	89.2	347	2	Q5AQN3	EMENI
6	288	89.2	350	2	Q8TFUB	EMENI
7	286	88.5	429	2	Q4HTT5	GIBZE
8	285	88.2	433	2	Q4WEY8	ASPFU
9	178	55.1	260	2	Q6BQC2	DEBHA
10	175	54.3	273	2	Q6CKQ1	KLJUL
11	175	54.3	299	2	Q6CEV1	YARLI
12	175	54.2	230	1	HAC1	YEAST
13	174	53.9	357	2	Q5AAE2	CANAL
14	168	52.0	228	2	Q75BQ5	ASHGO
15	148	45.8	329	2	Q6FLY3	CANGA
16	107	33.1	336	2	Q5VR11	ORYSA
17	106	32.8	168	1	HY5	ARATH
18	102	31.6	69	2	Q58L60	BRACM
19	102	31.6	109	2	Q84XX8	ERARP
20	101	31.3	158	1	HY5	LYCES
21	100.5	31.1	141	2	Q677A7	9ASPA
22	98	30.3	176	2	Q69XK6	ORYSA
23	96	29.7	252	2	Q6NX18	XENTR
24	94	29.1	188	2	Q6ZHT8	ORYSA
25	94	29.1	321	2	Q6GRY7	LOTJA
26	94	29.1	646	2	Q6AU90	ORYSA
27	94	29.1	686	2	Q61D33	CAENR
28	93	28.8	120	2	Q682B6	ARATH
29	93	28.8	149	1	HYH	ARATH
30	92	28.5	350	2	Q90ZR7	XENLA
31	92	28.5	396	2	Q7ZYC2	XENLA

32	92	28.5	460	2	Q93XA0	PHAVU
33	91	28.2	322	2	Q39896	glycine max
34	91	28.2	326	2	Q39895	glycine max
35	90.5	28.0	690	2	Q44743	caenorhabdi
36	90	27.9	404	1	CRBB3	MOUSE
37	89	27.6	371	2	Q5TCV1	HUMAN
38	89	27.6	371	2	Q5R5Z9	PONPY
39	89	27.6	395	1	CRBB3	HUMAN
40	88.5	27.4	208	2	Q5DFK2	SCHJA
41	88.5	27.4	437	2	Q569T3	XENLA
42	88	27.2	322	2	Q04234	VICFA
43	87.5	27.1	176	2	Q98RX1	GUILLAR
44	87.5	27.1	585	2	Q4P8Q4	USTILAGO
45	87.5	27.1	834	2	Q54Y73	DICDI

ALIGNMENTS

RESULT 1
Q8TFP3 TRIRE
ID Q8TFP3 TRIRE PRELIMINARY; PRT; 451 AA.
AC Q8TFP3;
DT 01-JUN-2002 (Tremblrel. 21, Created)
DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)
DT 01-MAR-2004 (Tremblrel. 26, Last annotation update)
DE Transcription factor.
GN Name=hacl;
OS Trichoderma reesei (Hypocrea jecorina).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Hypocreomycetidae; Hypocreales; Hypocreaceae; Hypocrea.
OX NCBI_TaxID=51453;
RN [1]
NUCLEOTIDE SEQUENCE.
RA Salohelmo M.L.A., Valkonen M., Penttilae M.E.;
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; A7413272; CAC8374.1; --; Genomic_DNA.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR011700; bZIP 2.
DR InterPro; IPR004827; TF_bZIP.
DR Pfam; PF07716; bZIP_2; 1.
DR SMART; SM00338; BRLZ; 1.
DR PROSITE; PS0217; bZIP; 1.
KW DNA-binding; Nuclear protein.
SQ SEQUENCE 451 AA; 49277 MW; 600F10B471EA3AD3 CRC64;

Query Match 100.0%; Score 323; DB 2; Length 451;
Best Local Similarity 100.0%; Pred. No. 5.1e-21;
Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	EKKPVKRSWGQVLPEPKTLP	PPRKRAKTEKEQRRVERVLNRRAAQSRRKRLEV	60
DB	84	EKKPVKRSWGQVLPEPKTLP	PPRKRAKTEKEQRRVERVLNRRAAQSRRKRLEV	143
QY	61	EALE 64		
DB	144	EALE 147		

RESULT 2
Q7SHF0 NEUCR
ID Q7SHF0 NEUCR PRELIMINARY; PRT; 430 AA.
AC Q7SHF0;
DT 01-MAR-2004 (Tremblrel. 26, Created)
DT 01-MAR-2004 (Tremblrel. 26, Last sequence update)
DT 01-MAR-2004 (Tremblrel. 26, Last annotation update)
DE Predicted protein.
GN Name=NCU01856.1;
OS Neurospora crassa.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.

```

OX NCBI_TaxID=5141;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=0R74A;
RA Galagan J.E., Calvo S.E., Borkovich K.A., Selker E.U., Read N.D.,
RA Jaffe D., Fitzhugh W., Ma L.-J., Smirnov S., Purcell S., Rehman B.,
RA Elkins T., Engels R., Wang S., Nielsen C.B., Butler J., Endrizzi M.,
RA Qui D., Ianakiev P., Pedersen D., Nelson M., Washburne M.,
RA Selltreinikoff C.P., Kinsey J.A., Braun E.L., Zelter A., Schulte U.,
RA Kothe G.O., Jedd G., Newes W., Staben C., Marcotte E., Greenberg D.,
RA Roy A., Foley K., Naylor J., Thomann N., Barrett R., Gnerre S.,
RA Kamal M., Kamysseelis M., Mauceli E., Bielke C., Rudd S., Frishman D.,
RA Krystofova S., Rasmussen C., Metzenberg R.L., Perkins D.D., Kroken S.,
RA Cogoni C., Macino G., Catcheside D., Li W., Pratt R.J., Omani S.A.,
RA DeSouza C.C., Glass L., Orbach M.J., Berglund J., Voelker R.,
RA Yarden O., Pflamann M., Seiler S., Dunlap J., Radford A., Aramayo R.,
RA Natvig D.O., Alex L.A., Mannhaupt G., Eboole D.J., Freitag M.,
RA Paulsen I., Sachs M.S., Lander E.S., Nussbaum C., Birren B.;
RT "The Genome Sequence of the Filamentous Fungus Neurospora crassa.";
RL Nature 0:0-0(2003).
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC EMBL; AABX01000004; EAA36251.1; -; Genomic_DNA.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR011700; bZIP_2.
DR InterPro; IPR004827; TF_bZIP.
DR Pfam; PF07716; bZIP_2; 1.
DR PROSITE; PS50217; bZIP; 1.
SQ SEQUENCE 430 AA; 46599 MW; 0EE0657CFA6160DA CRC64;

Query Match 98.4%; Score 319; DB 2; Length 430;
Best Local Similarity 98.4%; Pred. No. 1.1e-20;
Matches 63; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 EKKPVKKRSGQVLPEPTNLPPRKRAKTEDEKEQRRVERVLNRRAAQSRRKRRLV 60
Db 87 EKKPVKKRSGQVLPEPTNLPPRKRAKTEDEKEQRRVERVLNRRAAQSRRKRRLV 146

QY 61 EALE 64
Db 147 EGLE 150

RESULT 3
Q51KW8 MAGGR
ID Q51KW8_MAGGR PRELIMINARY; PRT; 556 AA.
AC Q51KW8;
DT 13-SEP-2005 (TrEMBLrel. 31, Created)
DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
DE Hypothetical protein.
OS ORFNames=MG09010.4;
GN Magnaporthe grisea 70-15.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariomycetes incertae sedis; Magnaporthaceae; Magnaporthe.
OX NCBI_TaxID=242507;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=70-15;
RA Birren B., Nuebaum C., Abebe A., Abouelleil A., Adekoya E.,
RA Ait-zahra M., Allen N., Allen T., An P., Anderson M., Anderson S.,
RA Arachchi H., Armbruster J., Bachantsang P., Baldwin J., Barry A.,
RA Bayul T., Blithsteyn B., Bloom T., Blye J., Boguslavskiy L.,
RA Borowsky M., Boukhgalter B., Brunache A., Butler J., Calixte N.,
RA Calvo S., Camarata J., Campo K., Chang J., Cheshatsang Y., Citroen M.,
RA Collimore A., Considine T., Cook A., Cooke P., Corum B., Cuomo C.,
RA David R., Dawoe T., Degray S., Dodge S., Dooley K., Dorje P.,
RA Dorjee K., Dorris L., Duffey N., Dupes A., Elkins T., Engels R.,
RA Erickson J., Farina A., Faro S., Ferreira P., Fischer H.,
RA Fitzgerald M., Foley K., Gage D., Galagan J., Gearin G., Gnerre S.,

```

```

RA Gnrirke A., Goyette A., Graham J., Grandbois E., Gyaltsen K., Hafez N.,
RA Hagopian D., Hagos B., Hall J., Hatcher B., Heller A., Higgins H.,
RA Honan T., Horn A., Houde N., Hughes L., Hulme W., Huseby E., Iliev I.,
RA Jaffe D., Jones C., Kamal M., Kamat A., Kamysseelis M., Karlsson E.,
RA Kells C., Kieu A., Kishner P., Kodira C., Kulbokas E., Labutti K.,
RA Lama D., Landers T., Leger J., Levine S., Lewis D., Lewis T.,
RA Lindblad-toh K., Liu X., Lokyitsang T., Lokyitsang Y., Lucien O.,
RA Lui A., Ma L.J., Mabbitt R., Macdonald J., Maclean C., Major J.,
RA Manning J., Marabella R., Maru K., Matthews C., Mauceli E.,
RA McCarthy M., McDonough S., Mcghee T., Meldrim J., Meneus L.,
RA Mesirov J., Mihalev A., Mihova T., Mikkelsen T., Mlenga V., Moru K.,
RA Mozes J., Mulrain L., Munson G., Naylor J., Neves C., Nguyen C.,
RA Nguyen N., Nguyen T., Nicol R., Nielsen C., Nizzari M., Norbu C.,
RA Norbu N., O'donnell P., Okawo O., O'leary S., Omotosho B.,
RA O'Neill K., Osman S., Parker S., Perrin D., Phunkhang P., Pignani B.,
RA Purcell S., Rachupka T., Ramasamy U., Rameau R., Ray V., Raymond C.,
RA Retta R., Richardson S., Rise C., Rodriguez J., Rogers J., Rogov P.,
RA Rutnan M., Schupbach R., Seaman C., Settipalli S., Sharpe T.,
RA Sheridan J., Sherpa N., Shi J., Smirnov S., Smith C., Sougnez C.,
RA Spencer B., Stalker J., Stange-thomann N., Stavropoulos S.,
RA Stetson K., Stone C., Stone S., Stubbs M., Talamas J., Tchuinga P.,
RA Tensley P., Tesfaye S., Theodore J., Thoulutsang Y., Topham K.,
RA Towey S., Tsamia T., Tsomo N., Vallee D., Vassiliev H.,
RA Venkataraman V., Vinson J., Vo A., Wade C., Wang S., Wangchuk T.,
RA Wangdi T., Whittaker C., Wilkinson J., Wu Y., Wyman D., Yadav S.,
RA Yang S., Yang X., Yeager S., Yee E., Young G., Zainoun J., Zembeck L.,
RA Zimmer A., Zody M., Lander E.;
RT "The genome sequence of Magnaporthe grisea.";
RL Submitted (OCT-2003) to the EMBL/GenBank/DBJ databases.
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=70-15;
RA Dean R., Mitchell T., Brown D., Pan H., Thon M.;
RL Submitted (OCT-2003) to the EMBL/GenBank/DBJ databases.
RN [3]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=70-15;
RA Zhu H., Blackmon B.;
RL Submitted (OCT-2003) to the EMBL/GenBank/DBJ databases.
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC EMBL; AACU01001606; EAA47880.1; -; Genomic_DNA.
DR InterPro; IPR011700; bZIP_2.
DR InterPro; IPR004827; TF_bZIP.
DR Pfam; PF07716; bZIP_2; 1.
DR SMART; SM00338; BRLZ; 1.
DR PROSITE; PS50217; bZIP; 1.
KW DNA-binding; Hypothetical protein; Nuclear protein.
SQ SEQUENCE 556 AA; 60798 MW; CFFDC4C8D2DC4963C CRC64;

Query Match 91.6%; Score 296; DB 2; Length 556;
Best Local Similarity 93.7%; Pred. No. 1.7e-18;
Matches 59; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 2 KKPVKRKSQGVLPPEPTNLPPRKRAKTEDEKEQRRVERVLNRRAAQSRRKRLEYE 61
Db 88 EKKTKRKSQGVLPPEPTNLPPRKRAKTEDEKEQRRVERVLNRRAAQSRRKRQVE 147

QY 62 ALE 64
Db 148 ALE 150

RESULT 4
Q6W8X2 ASPNG
ID Q6W8X2 ASPNG PRELIMINARY; PRT; 342 AA.
AC Q6W8X2;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Transcription factor HACA.
GN Name=haca;

```

```
OS Aspergillus niger.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
OX NCBI_TaxID=5061;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Mulder H.J.;
RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY303684; AAQ73495.1; -; Genomic_DNA.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR011700; bZIP_2.
DR InterPro; IPR004827; TF_bZIP.
DR Pfam; PF07716; bZIP_2; 1.
DR SMART; SM00338; BRLZ; 1.
DR PROSITE; PS0217; bZIP; 1.
DR PROSITE; PS00036; bZIP_BASIC; UNKNOWN_1.
KW DNA-binding; Nuclear protein.
SQ SEQUENCE 342 AA; 37148 MW; 623B3941A55C9C05 CRC64;

Query Match 90.4%; Score 292; DB 2; Length 342;
Best Local Similarity 89.1%; Pred. No. 2.4e-18;
Matches 57; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 EKKPVKRSWGQVLPEPKTNLPKRKATDEKEQRRVERVLRNRAAQSRRKRLEV 60
DB 46 EKKPVKRSWGQELVPVKTNLPKRKATDEKEQRRVERVLRNRAAQSRRKRLEV 105
QY 61 EALE 64
DB 106 EXLE 109

RESULT 5
QSAQN3 EMENI
ID QSAQN3_EMENI PRELIMINARY; PRT; 347 AA.
AC QSAQN3;
DT 10-MAY-2005 (TrEMBLrel. 30, Created)
DT 10-MAY-2005 (TrEMBLrel. 30, Last sequence update)
DE Hypothetical protein.
GN ORFNames=AN9397.2;
OS Aspergillus nidulans FGSC A4.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; Emericella.
OX NCBI_TaxID=227321;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Birren B., Nusbaum C., Abouelleil A., Allen N., Anderson S.,
RA Arachchi H.M., Barna N., Bastien V., Bloom T., Boguslavskiy L.,
RA Boukhgalter B., Butler J., Calvo S.E., Camarata J., Chang J.,
RA Choepel Y., Collymore A., Cook A., Cooke P., Corum B., Dearrellano K.,
RA Diaz J.S., Dodge S., Dooley K., Dorris L., Elkins T., Engels R.,
RA Erickson J., Faro S., Ferreira P., FitzGerald M., Gage D., Galagan J.,
RA Gardyna S., Gnirre S., Graham L., Grand-Pierre N., Hafez N.,
RA Hagopian D., Hagos B., Hall J., Horton L., Hulme W., Iliev I.,
RA Jaffe D., Johnson R., Jones C., Kanat M., Kanat A., Karatas A.,
RA Kells C., Landers T., Levine R., Lindblad-Toh K., Liu G., Lui A.,
RA Ma L.-J., Mabbitt R., MacLean C., Macdonald P., Major J., Manning J.,
RA Matthews C., Mauceli E., McCarthy M., Meldrim J., Meneus L.,
RA Mihova T., Mlenga V., Murphy T., Naylor J., Nguyen C., Nicol R.,
RA Nielsen C.B., Norbu C., O'Connor T., O'Donnell P., O'Neill D.,
RA Oliver J., Peterson K., Phunkhang P., Pierre N., Purcell S.,
RA Rachupka A., Ramasamy U., Raymond C., Retta R., Rise C., Rogov P.,
RA Roman J., Schauer S., Schupack R., Seaman S., Severy P., Smirnov S.,
RA Smith C., Spencer B., Stange-Thomann N., Stojanovic N., Stubbs M.,
RA Talamas J., Tesfaye S., Theodore J., Topham K., Travers M.,
RA Vassiliev H., Venkataraman V.S., Viel R., Vo A., Wang S., Wilson B.,
RA Wu X., Wyman D., Young G., Zainoun J., Zembek L., Zimmer A., Zody M.,
RA Lander E.;
RT "Genome Sequence of Aspergillus nidulans.";
```

```
RL Submitted (JAN-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; AACD01000172; EAA6464.1; -; Genomic_DNA.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
KW DNA-binding; Hypothetical protein; Nuclear protein.
SQ SEQUENCE 347 AA; 37841 MW; A7CAAFEE761B3E1 CRC64;

Query Match 89.2%; Score 288; DB 2; Length 347;
Best Local Similarity 87.5%; Pred. No. 5.5e-18;
Matches 56; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 1 EKKPVKRSWGQVLPEPKTNLPKRKATDEKEQRRVERVLRNRAAQSRRKRLEV 60
DB 53 EKKPAKRKSWGQELVPVKTNLPKRKATDEKEQRRVERVLRNRAAQSRRKRLEV 112
QY 61 EALE 64
DB 113 EXLE 116

RESULT 6
Q8TFU8 EMENI
ID Q8TFU8_EMENI PRELIMINARY; PRT; 350 AA.
AC Q8TFU8;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Transcription factor.
GN Name=hacA;
OS Emericella nidulans (Aspergillus nidulans).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; Emericella.
OX NCBI_TaxID=162425;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Saloheimo M.L.A., Valkonen M., Penttilae M.E.;
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ413273; CAC88375.1; -; Genomic_DNA.
DR HSSP; P05412; 1JNN.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR011700; bZIP_2.
DR InterPro; IPR004827; TF_bZIP.
DR Pfam; PF07716; bZIP_2; 1.
DR SMART; SM00338; BRLZ; 1.
DR PROSITE; PS0217; bZIP; 1.
DR PROSITE; PS00036; bZIP_BASIC; UNKNOWN_1.
KW DNA-binding; Nuclear protein.
SQ SEQUENCE 350 AA; 38192 MW; 4F78340096064F44 CRC64;

Query Match 89.2%; Score 288; DB 2; Length 350;
Best Local Similarity 87.5%; Pred. No. 5.6e-18;
Matches 56; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 1 EKKPVKRSWGQVLPEPKTNLPKRKATDEKEQRRVERVLRNRAAQSRRKRLEV 60
DB 53 EKKPAKRKSWGQELVPVKTNLPKRKATDEKEQRRVERVLRNRAAQSRRKRLEV 112
QY 61 EALE 64
DB 113 EXLE 116

RESULT 7
Q4HTT5 GIBZE
ID Q4HTT5_GIBZE PRELIMINARY; PRT; 429 AA.
AC Q4HTT5;
DT 13-SEP-2005 (TrEMBLrel. 31, Created)
DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
DE Hypothetical protein.
```


RA Bouchier C., Caudron B., Scarpelli C., Gaillardin C., Weissenbach J.,
 RT "Genome evolution in yeasts."
 RL Nature 430:35-44 (2004).
 DR EMBL; CR382137; CAG87828.1; -; Genomic_DNA.
 DR GO; GO:0005634; C:nucleus; IEA.
 DR GO; GO:0003677; F:DNA binding; IEA.
 DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
 DR InterPro; IPR011700; bZIP_2.
 DR Pfam; PF07716; bZIP_2; 1.
 DR SMART; SM00338; BRLZ; 1.
 DR PROSITE; PS00217; bZIP; 1.
 DR PROSITE; PS00036; bZIP_BASIC; UNKNOWN_1.
 KW Complete proteome; DNA-binding; Nuclear protein.
 SQ SEQUENCE 260 AA; 29442 MW; 49B1819E393BFF6 CRC64;

Query Match 55.1%; Score 178; DB 2; Length 260;
 Best Local Similarity 76.1%; Pred. No. 3.2e-08;
 Matches 35; Conservative 7; Mismatches 4; Indels 0; Gaps 0;

QY 19 KTNLPKRAKTEDEKEQRRVERVLNRRRAAQSRRERKLEVEALE 64
 :::
 DB 14 KSSLPPKRAKTEKEQRRVERILNRRRAAHASREKKRHVEYLE 59

RESULT 10
 ID Q6CKQ1_KLUJA PRELIMINARY; PRT; 273 AA.
 AC Q6CKQ1;
 DT 25-OCT-2004 (TrEMBLrel. 28, Created)
 DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
 DE Kluyveromyces lactis strain NRRL Y-1140 chromosome F of strain NRRL Y-1140 of Kluyveromyces lactis.
 DE 1140 of Kluyveromyces lactis.
 GN OrderedLocusNames=KLLA0F08976g;
 OS Kluyveromyces lactis (Yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Kluyveromyces.
 OX NCBI_TaxID=28985;
 RN [1]
 RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
 RC STRAIN=CLIB 122 / E 150;
 RX PubMed=15229592; DOI=10.1038/nature02579;
 RA Lafontaine I., de Montigny J., March C., Neuveglise C., Talla E.,
 RA Goffard N., Frangoul L., Aigle M., Anthouard V., Babour A., Barbe V.,
 RA Barnay S., Blanchin S., Beckerich J.-M., Beyne E., Bleykaesten C.,
 RA Boisrame A., Boyer J., Cattolico L., Confanier F., de Daruvar A.,
 RA Despons L., Fabre E., Fairhead C., Ferry-Dumazet H., Groppi A.,
 RA Hantraye F., Hennequin C., Jauniaux N., Joyet P., Kachouri R.,
 RA Kerrest A., Koszul R., Lemaire M., Lesur I., Ma L., Muller H.,
 RA Nicaud J.-M., Nikolski M., Oztas S., Ozier-Kalogeropoulos O.,
 RA Pellenz S., Potier S., Richard G.-P., Straub M.-L., Suleau A.,
 RA Swennen D., Tekai F., Wesolowski-Louvel M., Westhof E., Wirth B.,
 RA Zentou-Meyer M., Zivanovic Y., Bolotin-Fukuhara M., Thierry A.,
 RA Bouchier C., Caudron B., Scarpelli C., Gaillardin C., Weissenbach J.,
 RA Winkler P., Souciet J.-L.;
 RT "Genome evolution in yeasts."
 RL Nature 430:35-44 (2004).
 RC STRAIN=CBS 2359 / IFO 1267 / NRRL Y-1140 / WM37;
 RX PubMed=15229592; DOI=10.1038/nature02579;
 RA Lafontaine I., de Montigny J., March C., Neuveglise C., Talla E.,
 RA Goffard N., Frangoul L., Aigle M., Anthouard V., Babour A., Barbe V.,
 RA Barnay S., Blanchin S., Beckerich J.-M., Beyne E., Bleykaesten C.,
 RA Boisrame A., Boyer J., Cattolico L., Confanier F., de Daruvar A.,
 RA Despons L., Fabre E., Fairhead C., Ferry-Dumazet H., Groppi A.,
 RA Hantraye F., Hennequin C., Jauniaux N., Joyet P., Kachouri R.,
 RA Kerrest A., Koszul R., Lemaire M., Lesur I., Ma L., Muller H.,
 RA Nicaud J.-M., Nikolski M., Oztas S., Ozier-Kalogeropoulos O.,
 RA Pellenz S., Potier S., Richard G.-P., Straub M.-L., Suleau A.,
 RA Swennen D., Tekai F., Wesolowski-Louvel M., Westhof E., Wirth B.,
 RA Zentou-Meyer M., Zivanovic Y., Bolotin-Fukuhara M., Thierry A.,
 RA Bouchier C., Caudron B., Scarpelli C., Gaillardin C., Weissenbach J.,
 RA Winkler P., Souciet J.-L.;
 RT "Genome evolution in yeasts."
 RL Nature 430:35-44 (2004).
 DR EMBL; CR382126; CAG98196.1; -; Genomic_DNA.
 DR GO; GO:0005634; C:nucleus; IEA.
 DR GO; GO:0003677; F:DNA binding; IEA.
 DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
 DR InterPro; IPR011616; bZIP_1.
 DR InterPro; IPR004827; TF_bZIP.
 DR Pfam; PF00170; bZIP_1; 1.
 DR SMART; SM00338; BRLZ; 1.
 DR PROSITE; PS00036; bZIP_BASIC; 1.
 KW Complete proteome; DNA-binding; Nuclear protein.
 SQ SEQUENCE 273 AA; 30429 MW; 97DFD7B1F8CAF5 CRC64;

Query Match 55.1%; Score 178; DB 2; Length 273;

Best Local Similarity 73.9%; Pred. No. 3.4e-08;
 Matches 34; Conservative 6; Mismatches 6; Indels 0; Gaps 0;

QY 19 KTNLPKRAKTEDEKEQRRVERVLNRRRAAQSRRERKLEVEALE 64
 :::
 DB 15 KPTLPKRAKTEKEQRRVERILNRRRAAQAHSREKKLHVORLE 60

RESULT 11
 Q6CEV1_YARLI PRELIMINARY; PRT; 299 AA.
 AC Q6CEV1;
 DT 25-OCT-2004 (TrEMBLrel. 28, Created)
 DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
 DE Yarrowia lipolytica chromosome B of strain CLIB99 of Yarrowia lipolytica.
 GN OrderedLocusNames=YALI0B12716g;
 OS Yarrowia lipolytica (Candida lipolytica).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Dipodascaceae; Yarrowia.
 OX NCBI_TaxID=4952;
 RN [1]
 RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
 RC STRAIN=CLIB 122 / E 150;
 RX PubMed=15229592; DOI=10.1038/nature02579;
 RA Dujon B., Sherman D., Fischer G., Durrens P., Casaregola S.,
 RA Lafontaine I., de Montigny J., March C., Neuveglise C., Talla E.,
 RA Goffard N., Frangoul L., Aigle M., Anthouard V., Babour A., Barbe V.,
 RA Barnay S., Blanchin S., Beckerich J.-M., Beyne E., Bleykaesten C.,
 RA Boisrame A., Boyer J., Cattolico L., Confanier F., de Daruvar A.,
 RA Despons L., Fabre E., Fairhead C., Ferry-Dumazet H., Groppi A.,
 RA Hantraye F., Hennequin C., Jauniaux N., Joyet P., Kachouri R.,
 RA Kerrest A., Koszul R., Lemaire M., Lesur I., Ma L., Muller H.,
 RA Nicaud J.-M., Nikolski M., Oztas S., Ozier-Kalogeropoulos O.,
 RA Pellenz S., Potier S., Richard G.-P., Straub M.-L., Suleau A.,
 RA Swennen D., Tekai F., Wesolowski-Louvel M., Westhof E., Wirth B.,
 RA Zentou-Meyer M., Zivanovic Y., Bolotin-Fukuhara M., Thierry A.,
 RA Bouchier C., Caudron B., Scarpelli C., Gaillardin C., Weissenbach J.,
 RA Winkler P., Souciet J.-L.;
 RT "Genome evolution in yeasts."
 RL Nature 430:35-44 (2004).
 DR EMBL; CR382128; CAG83062.1; -; Genomic_DNA.
 DR GO; GO:0005634; C:nucleus; IEA.
 DR GO; GO:0003677; F:DNA binding; IEA.
 DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
 DR InterPro; IPR011700; bZIP_2.
 DR InterPro; IPR004827; TF_bZIP.
 DR Pfam; PF07716; bZIP_2; 1.
 DR SMART; SM00338; BRLZ; 1.
 DR PROSITE; PS00217; bZIP; 1.
 DR PROSITE; PS00036; bZIP_BASIC; UNKNOWN_1.
 KW Complete proteome; DNA-binding; Nuclear protein.
 SQ SEQUENCE 299 AA; 32817 MW; F536110CFB23ABB2 CRC64;

Query Match 54.3%; Score 175.5; DB 2; Length 299;
 Best Local Similarity 63.8%; Pred. No. 6.1e-08;
 Matches 37; Conservative 8; Mismatches 12; Indels 1; Gaps 1;

QY 7 KRKSGQVLPEPKTNLPKRAKTEDEKEQRRVERVLNRRRAAQSRRERKLEVEALE 64
 :::
 DB 33 KRKK-DLTLPLPAGALPPKRAKTEKEQRRVERILNRRRAAQAHSREKKRHLEBLE 89

RESULT 12
 HACL_YEAST STANDARD; PRT; 230 AA.
 ID HACL_YEAST
 AC P41546; P87040;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 13-SEP-2005 (Rel. 48, Last annotation update)
 DE HACL protein
 GN Name=HACL1; Synonyms=ERN4, IRE2; OrderedLocusNames=YFL031W;


```

RA Dungan J., Kuo A., Newport G., Lan C.-Y., Iijima C., Adegbola O.,
RA Roberts J., Persson K., Donnelly S., Favoreto S., Trung K.-W.,
RA Jones T., Scherer S., Agabian N.;
RT "Annotatation of the Genome of Candida albicans.";
RL Submitted (APR-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; AACQ01000041; EAK99529.1; -; Genomic DNA.
DR EMBL; AACQ01000040; EAK99617.1; -; Genomic DNA.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003677; P:DNA binding; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
KW DNA-binding; Hypothetical protein; Nuclear protein.
SQ SEQUENCE 357 AA; 40127 MW; 8E1C633A2B4ECDAE CRC64;

Query Match 53.9%; Score 174; DB 2; Length 357;
Best Local Similarity 71.7%; Pred. No. 9.9e-08;
Matches 33; Conservative 8; Mismatches 5; Indels 0; Gaps 0;

QY 19 KTNLPKRKAKTEDEKEQRRVRLNRRRAAQSRRKRKLEVEALE 64
Db 54 KSTLPPRKRAKTEQEEQKRIELNRRRAAHASREKKRGHVEYLE 99

RESULT 14
Q75BQ5 ASHGO
ID Q75BQ5 ASHGO PRELIMINARY; PRT; 228 AA.
AC Q75BQ5
DT 05-JUL-2004 (TREMBlrel. 27, Created)
DT 05-JUL-2004 (TREMBlrel. 27, Last sequence update)
DT 05-JUL-2004 (TREMBlrel. 27, Last annotation update)
DE ACR216Cp.
GN Name=ACR216C;
OS Ashbya gossypii (Eremothecium gossypii).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Eremothecium.
OX NCBI_TaxID=33169;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=ATCC 10895; DOI=10.1126/science.1095781;
RX PubMed=15001715;
RA Dietrich F.S., Voegel S., Brachat S., Lerch A., Gates K., Steiner S.,
RA Mohr C., Poehlmann R., Luedi P., Choi S., Wing R.A., Flavier A.,
RA Gaffney T.D., Philippsen P.;
RT "The Ashbya gossypii genome as a tool for mapping the ancient
RT Saccharomyces cerevisiae genome.";
RL Science 304:304-307(2004).
DR EMBL; AB016816; AAS51442.1; -; Genomic DNA.
DR AGD; ACR216C; -.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR011700; bZIP 2.
DR InterPro; IPR004827; TF_bZIP.
DR Pfam; PF07716; bZIP 2; 1.
DR PROSITE; PS00036; bZIP_BASIC; 1.
KW Complete proteome.
SQ SEQUENCE 228 AA; 24914 MW; 605A60D78633F038 CRC64;

Query Match 52.0%; Score 168; DB 2; Length 228;
Best Local Similarity 69.6%; Pred. No. 2.2e-07;
Matches 32; Conservative 8; Mismatches 6; Indels 0; Gaps 0;

QY 19 KTNLPKRKAKTEDEKEQRRVRLNRRRAAQSRRKRKLEVEALE 64
Db 14 KSTLPPRKRAKTEQEEQRRRIELNRRRAAHQSRKKGLHLYLE 59

RESULT 15
Q6FLY3 CANGA
ID Q6FLY3 CANGA PRELIMINARY; PRT; 329 AA.
AC Q6FLY3
DT 05-JUL-2004 (TREMBlrel. 27, Created)
DT 05-JUL-2004 (TREMBlrel. 27, Last sequence update)
DT 05-JUL-2004 (TREMBlrel. 27, Last annotation update)

```

```

DE Candida glabrata strain CBS138 chromosome K complete sequence.
GN OrderedLocusNames=CAGL0K12540g;
OS Candida glabrata (yeast) (Torulopsis glabrata).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX NCBI_TaxID=5478;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=ATCC 2001 / CBS 138;
RX PubMed=15229592; DOI=10.1038/nature02579;
RA Dufon B., Sherman D., Fischer G., Durrans P., Casaregola S.,
RA Lafontaine I., de Montigny J., Marck C., Neuveglise C., Talla E.,
RA Goffard N., Frangeul L., Aigle M., Anthouard V., Babour A., Barbe V.,
RA Barnay S., Blanchin S., Beckerich J.-M., Beyne E., Bleykasten C.,
RA Boismame A., Boyer J., Cattolico L., Confanier F., de Daruvar A.,
RA Despons L., Fabre E., Fairhead C., Ferry-Dumazet H., Groppi A.,
RA Hantraye F., Hennequin C., Jauniaux N., Joyet P., Kachouri R.,
RA Kerrest A., Koszul R., Lemaire M., Lesur I., Ma L., Muller H.,
RA Nicaud J.-M., Nikoleki M., Oztae S., Ozier-Kalogeropoulos O.,
RA Pellenz S., Potier S., Richard G.-F., Straub M.-L., Suleau A.,
RA Swennen D., Tekala F., Wesolowski-Louvel M., Westhof E., Wirth B.,
RA Zeniou-Meyer M., Zivanovic Y., Bolotin-Fukuhara M., Thierry A.,
RA Bouchier C., Caudron B., Scarpelli C., Gaillardin C., Weissenbach J.,
RA Wincker P., Souciet J.-L.;
RT "Genome evolution in yeasts.";
RL Nature 430:35-44(2004).
DR EMBL; CR380957; CAG61724.1; -; Genomic DNA.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR011616; bZIP 1.
DR InterPro; IPR004827; TF_bZIP.
DR Pfam; PF00170; bZIP 1; 1.
DR SMART; SM00338; BRLZ; 1.
DR PROSITE; PS0217; bZIP; 1.
DR PROSITE; PS00036; bZIP_BASIC; 1.
KW Complete proteome; DNA-binding; Nuclear protein.
SQ SEQUENCE 329 AA; 37162 MW; 469614BF3AF2282 CRC64;

Query Match 45.8%; Score 148; DB 2; Length 329;
Best Local Similarity 63.0%; Pred. No. 2e-05;
Matches 29; Conservative 10; Mismatches 7; Indels 0; Gaps 0;

QY 19 KTNLPKRKAKTEDEKEQRRVRLNRRRAAQSRRKRKLEVEALE 64
Db 26 ETWMPKRKRAKTEKESKIRKIQLRNKRAKQSKRDKRNYVANLE 71

Search completed: November 22, 2005, 15:48:41
Job time : 154.5 secs

```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 23, 2005, 03:04:00 ; Search time 26.5576 Seconds
(without alignments)
1125.017 Million cell updates/sec

Title: US-10-663-450-60

Perfect score: 342

Sequence: 1 KSTLPKRAKTKKEQR.....CSLENNLNSVNEKLDHHE 68

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_21.*

1: Geneseqp1980s.*

2: Geneseqp1990s.*

3: Geneseqp2000s.*

4: Geneseqp2001s.*

5: Geneseqp2002s.*

6: Geneseqp2003as.*

7: Geneseqp2003bs.*

8: Geneseqp2004s.*

9: Geneseqp2005s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	342	100.0	68	5	Aae15382 Yeast HAC
2	342	100.0	68	8	Adsi12859 Saccharom
3	342	100.0	84	5	Abp02534 Human ORF
4	342	100.0	200	8	Adsi43437 Bacterial
5	342	100.0	230	2	Aaw53806 Transcrip
6	342	100.0	230	8	Adt87049 Yeast Str
7	342	100.0	238	2	Aaw53807 Transcrip
8	196	57.3	450	4	Aab82975 Trichoder
9	196	57.3	451	5	Aae15371 Trichoder
10	196	57.3	451	8	Adsi12801 Trichoder
11	196	57.3	451	8	Adsi12804 Trichoder
12	178.5	52.2	409	7	Abos31144 A. thalia
13	178.5	52.2	409	7	Adbs31925 Plant (A.
14	178.5	52.2	409	8	Ado02271 Thalecres
15	176.5	51.6	342	4	Aab82977 Aspergill
16	176.5	51.6	342	5	Aae15381 Aspergill
17	176.5	51.6	342	8	Adsi12818 Aspergill
18	176.5	51.6	349	8	Adsi12805 Aspergill
19	176.5	51.6	349	8	Adsi12803 Aspergill
20	176.5	51.6	349	8	Adsi12863 Aspergill
21	176.5	51.6	350	4	Aab82976 Aspergill
22	176.5	51.6	350	5	Aae15372 Aspergill
23	176.5	51.6	386	5	Aae15379 Aspergill
24	176.5	51.6	386	8	Adsi12815 Aspergill

25	175	51.2	64	5	Aae15373 Trichoder
26	170	49.7	64	5	Aae15374 Aspergill
27	124.5	36.4	168	5	Aau93013 Arabidops
28	124.5	36.4	168	7	Add30174 Plant yie
29	124.5	36.4	168	8	Adi43893 Plant tra
30	124.5	36.4	211	3	Aag08861 Arabidops
31	121.5	35.5	192	4	Aab82614 Arabidops
32	121.5	35.5	192	4	Aab82615 Maize roo
33	121.5	35.5	192	4	Aab82616 Maize roo
34	121.5	35.5	192	4	Aag66525 Maize roo
35	121.5	35.5	192	4	Aag66526 Maize roo
36	116.5	34.1	170	8	Adm48147 Polypepti
37	115	33.6	23	3	Aay83623 Peptide f
38	114	33.3	185	9	Adm17162 Eucalyptu
39	112	32.7	163	9	Adw17580 Pinus rad
40	110	32.2	143	3	Aab33151 Pinus rad
41	107	31.3	120	3	Aag27808 Arabidops
42	107	31.3	135	3	Aag07181 Arabidops
43	107	31.3	149	3	Aag07180 Arabidops
44	107	31.3	149	3	Aag27807 Arabidops
45	107	31.3	188	3	Aag27806 Arabidops

ALIGNMENTS

RESULT 1

AAE15382

ID AAE15382 standard; protein; 68 AA.

AC AAE15382;

DT 07-MAR-2002 (first entry)

DE Yeast HAC1 protein DNA binding domain.

KW Heterologous protein secretion; unfolded protein response; UPR; lipase;
KW cellulase; carbohydrase; industry; purification; DNA binding domain;
KW HAC1 protein; Yeast.

OS Saccharomycetes cerevisiae.

PN US2001034045-A1.

PD 25-OCT-2001.

PF 23-MAR-2001; 2001US-00816277.

PR 24-MAR-2000; 2000US-00534692.

XX (GEMV) GENENCOR INT INC.

XX Penttila ME, Ward M, Wang H, Valkonen MJ, Saloheimo MLA;

XX WPI; 2002-033728/04.

XX Increasing secretion of heterologous proteins e.g. lipase and cellulase
in eukaryotic cells useful in industry to increase production and
facilitate purification, by inducing an elevated unfolded protein
response.

XX Example 3; Fig 10; 56pp; English.

XX The present invention relates to methods for increasing the secretion of
heterologous protein in eukaryotic cells by inducing an elevated unfolded
protein response (UPR). The method involves inducing the elevated UPR by
increasing the presence of proteins such as HAC1, HACA, PTC2 or IRE1 in
cells. The method and sequences are useful for increasing the secretion
of heterologous proteins (e.g. lipase, cellulase, carbohydrase) in
eukaryotic cells useful in industry to increase protein yields and to
facilitate purification. The present sequence is yeast HAC1 protein DNA
binding domain

```
SQ Sequence 68 AA;
Query Match 100.0%; Score 342; DB 5; Length 68;
Best Local Similarity 100.0%; Pred. No. 2e-31;
Matches 68; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 KSTLPPRKAKTKKEEKEQRRIERILNRRRAAHSREKRLHLQYLERKCSLLENLNSVN 60
Db 1 KSTLPPRKAKTKKEEKEQRRIERILNRRRAAHSREKRLHLQYLERKCSLLENLNSVN 60
QY 61 LEKLADHE 68
Db 61 LEKLADHE 68
RESULT 2
ADS12859
ID ADS12859 standard; protein; 68 AA.
XX
AC ADS12859;
DT 16-DEC-2004 (first entry)
XX
DE Saccharomyces cerevisiae hac1 DNA binding domain.
XX
KW unfolded protein response; UPR; HAC1; PTC2;
KW unfolded protein response modulator; enzyme production; hac1; yeast;
KW chaperone; foldase; DNA binding domain.
XX
OS Saccharomyces cerevisiae.
XX
PN US2004186070-A1.
XX
PD 23-SEP-2004.
XX
PF 15-SEP-2003; 2003US-00663450.
XX
PR 24-MAR-2000; 2000US-00534692.
PR 23-MAR-2001; 2001US-00816277.
XX
XX (GEMV ) GENENCOR INT INC.
PA
PI Penttila ME, Ward M, Wang H, Valkonen MJ, Saloheimo MLA;
XX WPI; 2004-707924/69.
XX
PT Increasing the secretion of a heterologous protein, such as a therapeutic
PT or an industrial enzyme, in genetically modified eukaryotic cells by
PT inducing an elevated unfolded protein response (UPR).
XX
PS Example 3; SEQ ID NO 60; 83pp; English.
XX
CC The invention describes a method of increasing the secretion of a
CC heterologous protein in a eukaryotic cell, comprising inducing an
CC elevated unfolded protein response (UPR). Also described are: an isolated
CC nucleic acid encoding a HAC1 protein, where the HAC1 protein induces
CC unfolded protein response and has less than 50% similarity to yeast HAC1
CC protein, an isolated nucleic acid encoding a HAC1 protein, where the HAC1
CC protein induces unfolded protein response and where the HAC1 protein
CC comprises a DNA binding region that has greater than 70% similarity to
CC the DNA binding region of filamentous fungi HAC1 protein; a protein
CC having unfolded protein response inducing activity and having greater
CC than 70% similarity to a fully defined amino acid sequence of 451, 349 or
CC 386, or 504 or 1232 amino acids (SEQ ID NO: 2, 4 or 16, or 12 or 14,
CC respectively) as given in the specification; a protein having an amino
CC acid sequence of SEQ ID NO: 2, 4, 8, 10, 12, 14 or 16; an isolated
CC nucleic acid encoding a PTC2 protein that modulates unfolded protein
CC response and has at least 70% similarity to a fully defined amino acid
CC sequence of 130 or 438 amino acids (SEQ ID NO: 8 or 10) as given in the
CC specification; an isolated nucleic acid encoding a IRE1 protein that
CC modulates unfolded protein response and has at least 60% similarity to a
CC fully defined amino acid sequence with SEQ ID NO: 12 or 14; a protein
CC having unfolded protein response modulating activity and having greater
```

```
than 70% similarity to SEQ ID NO: 8 or 10; and a cell containing a
CC heterologous nucleic acid encoding a protein having unfolded protein
CC response modulating activity and a heterologous nucleic acid encoding a
CC protein of interest to be secreted. The methods and compositions of
CC genetically manipulating cells to have an elevated unfolded protein
CC response (UPR) resulting in an increased capacity to produce secreted
CC proteins, are useful in e.g. production of therapeutic or industrial
CC enzymes. This is the amino acid sequence of Saccharomyces cerevisiae hac1
CC chaperone and foldase DNA binding domain.
XX
SQ Sequence 68 AA;
Query Match 100.0%; Score 342; DB 8; Length 68;
Best Local Similarity 100.0%; Pred. No. 2e-31;
Matches 68; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 KSTLPPRKAKTKKEEKEQRRIERILNRRRAAHSREKRLHLQYLERKCSLLENLNSVN 60
Db 1 KSTLPPRKAKTKKEEKEQRRIERILNRRRAAHSREKRLHLQYLERKCSLLENLNSVN 60
QY 61 LEKLADHE 68
Db 61 LEKLADHE 68
RESULT 3
ABP02534
ID ABP02534 standard; protein; 84 AA.
XX
AC ABP02534;
XX
DT 24-JUN-2002 (first entry)
XX
DE Human ORFX protein sequence SEQ ID NO:5050.
XX
KW Human; open reading frame; ORFX; gene therapy; cancer; cirrhosis;
KW hyperproliferative disorder; psoriasis; benign tumour; haemorrhage;
KW degenerative disorder; osteoarthritis; neurodegenerative disorder;
KW cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;
KW hypertension; hypothyroidism; cholesterol ester storage disease;
KW immune deficiency; immune disorder; infectious disease;
KW autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;
KW myasthenia gravis.
XX
OS Homo sapiens.
XX
PN WO200192523-A2.
XX
PD 06-DEC-2001.
XX
PF 29-MAY-2001; 2001WO-US010836.
XX
PR 30-MAY-2000; 2000US-0206132P.
PR 29-AUG-2000; 2000US-0228716P.
XX
XX (CURA-) CURAGEN CORP.
XX
PI Shinketsu RA, Leach MD;
XX WPI; 2002-106308/14.
DR N-PSDB; ABN18286.
XX
PT Novel human polypeptides and polynucleotides useful for diagnosing,
PT preventing and treating cardiovascular disease, neurodegenerative,
PT hyperproliferative disorders and autoimmune disorders.
PS Disclosure; SEQ ID NO 5050; 1037pp; English.
XX
CC The present invention describes substantially purified human proteins
CC (referred to as open reading frames, ORFX, where X is 1-11491 (see Table 1
CC in the specification). ABN15762 to ABN27252 encode the human ORFX
CC proteins given in ABP00010 to ABP11500. ORFX proteins are useful for
CC treating or preventing a pathology associated with an ORFX-associated
```

CC disorder in humans, and in the manufacture of a medicament for treating a
 CC syndrome associated with ORFX-associated disorder. ORFX polynucleotide
 CC sequences can be used in gene therapy. ORFX sequences can be used in the
 CC treatment of cancer, hyperproliferative disorders, cirrhosis of liver,
 CC psoriasis, benign tumours, keloid, degenerative disorders, haemorrhage,
 CC osteoarthritis, neurodegenerative disorders, disorders related to organ
 CC transplantation, cardiovascular diseases, diabetes mellitus, systemic
 CC lupus erythematosus, hypertension, hypothyroidism, cholesterol ester
 CC storage disease, various immune deficiencies and disorders, infectious
 CC diseases, autoimmune disorders such as multiple sclerosis, rheumatoid
 CC arthritis, autoimmune thyroiditis, myasthenia gravis, graft-versus-host
 CC disease and autoimmune inflammatory eye disease. ORFX proteins are also
 CC useful for treating burns, incisions, ulcers, for treating osteoporosis,
 CC bone degenerative disorders, or periodontal disease, and for gut
 CC protection or regeneration and treatment of lung or liver fibrosis,
 CC reperfusion injury in various tissues and conditions resulting from
 CC systemic cytokine damage. N.B. The sequence data for this patent did not
 CC form part of the printed specification, but was obtained in electronic
 CC format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences
 XX
 XX Sequence 84 AA;

Query Match 100.0%; Score 342; DB 5; Length 84;
 Best Local Similarity 100.0%; Pred. No. 2.5e-31;
 Matches 68; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KSTLPPRKRAKTKKEQRIERILNRRRAAHSREKKRLHLYLERKCSLLENLLNSVN 60
 DB 15 KSTLPPRKRAKTKKEQRIERILNRRRAAHSREKKRLHLYLERKCSLLENLLNSVN 74
 QY 61 LEKLADHE 68
 DB 75 LEKLADHE 82

RESULT 4
 ADS43437
 ID ADS43437 standard; protein; 200 AA.
 XX
 AC ADS43437;
 DT 02-DEC-2004 (first entry)
 XX Bacterial polypeptide #21867.
 DE
 XX Recombinant DNA construct; transformed plant; improved plant property;
 KW cold tolerance; heat tolerance; drought tolerance; herbicide; osmosis;
 KW pathogen tolerance; pest tolerance; plant disease resistance;
 KW cell cycle pathway modification; plant growth regulator;
 KW homologous recombination; seed oil yield; protein yield; carbohydrate;
 KW nitrogen; phosphorus; photosynthesis; lignin; galactomannan;
 KW bacterial polypeptide.
 XX Bacteria.

OS
 XX US2003233675-A1.
 PN
 XX 18-DEC-2003.
 XX
 XX 20-FEB-2003; 2003US-00369493.
 XX
 XX 21-FEB-2002; 2002US-0360039P.
 PR
 XX (CAOV/) CAO Y.
 PA (HINK/) HINKLE G J.
 PA (SLAT/) SLATER S C.
 PA (CHEN/) CHEN X.
 PA (GOLD/) GOLDMAN B S.
 XX
 XX Cao Y, Hinkle GJ, Slater SC, Chen X, Goldman BS;
 PI WPI; 2004-061375/06.
 XX
 DR
 XX

PT New recombinant DNA construct comprising a promoter positioned to provide
 PT for expression of a polynucleotide encoding a polypeptide from a
 PT microbial source, useful for producing plants with improved properties.
 XX
 PS Claim 1; SEQ ID NO 21867; 122pp; English.

XX The invention relates to a recombinant DNA construct comprising a
 CC promoter functional in a plant cell, where the promoter is positioned to
 CC provide for expression of a polynucleotide encoding a polypeptide from a
 CC microbial source. The invention also relates to a transformed plant
 CC comprising the recombinant DNA construct and a method of producing a
 CC transformed plant having an improved property. The plant is a crop plant
 CC such as maize or soybean. The method of producing a transformed plant
 CC having an improved property comprises transforming a plant with the
 CC recombinant DNA construct and growing the transformed plant, where the
 CC polynucleotide or polypeptide is useful for improving plant properties.
 CC The recombinant DNA construct is useful for producing plants with
 CC improved plant properties, e.g. improved cold, heat or drought tolerance,
 CC tolerance to herbicides, extreme osmotic conditions, pathogens or pests,
 CC increased resistance to plant disease, better growth rate by modification
 CC of the cell cycle pathway with plant growth regulators, increased rate of
 CC homologous recombination, modified seed oil or protein yield and/or
 CC content, improved yield by modification of carbohydrate, nitrogen or
 CC phosphorus use and/or uptake, by modification of photosynthesis or by
 CC providing improved plant growth and development under at least one stress
 CC condition, improved lignin production or improved galactomannan
 CC production. This sequence represents a bacterial polypeptide used in the
 CC scope of the invention. Note: The sequence data for this patent did not
 CC form part of the printed specification but was obtained in electronic
 CC format from USPTO at seqdata.uspto.gov/sequence.html.

XX Sequence 200 AA;

Query Match 100.0%; Score 342; DB 8; Length 200;
 Best Local Similarity 100.0%; Pred. No. 6.3e-31;
 Matches 68; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KSTLPPRKRAKTKKEQRIERILNRRRAAHSREKKRLHLYLERKCSLLENLLNSVN 60
 DB 23 KSTLPPRKRAKTKKEQRIERILNRRRAAHSREKKRLHLYLERKCSLLENLLNSVN 82
 QY 61 LEKLADHE 68
 DB 83 LEKLADHE 90

RESULT 5
 AAW53806
 ID AAW53806 standard; protein; 230 AA.
 XX
 AC AAW53806;
 XX
 DT 06-JUL-1998 (first entry)
 XX
 XX Transcription control protein ERN4.
 DE
 XX Transcription control protein; ERN4; chaperone protein; protein folding;
 KW transcription enzyme.
 KW
 XX Saccharomyces cerevisiae.
 OS
 XX JP10084970-A.
 PN
 XX 07-APR-1998.
 PD
 XX 31-OCT-1996; 96JP-00307085.
 PF
 XX 23-JUL-1996; 96JP-00213180.
 PR
 XX (HSPK-) HSP KENKYUSHO KK.
 XX
 XX WPI; 1998-264856/24.
 DR N-PSDB; AAV20067.
 DR

XX DNA encoding a yeast transcriptional control protein - useful in the
PT recombinant production of high yield of protein of interest.
XX
XX Claim 1; Page 14-15; 21pp; Japanese.
XX
CC This sequence is the yeast transcriptional control protein, designated
CC ERN4, of the invention. An alternatively spliced version of the protein
CC is shown in AA52806. Yeast cells transformed with a vector containing
CC the DNA can be used for the recombinant production of a protein of
CC interest at a high yield. The DNA controls a chaperone protein which in
CC turn is responsible for the correct folding of an enzyme involved in
CC transcription, hence the cells can be used to control the transcription
CC of a recombinant protein of interest
XX
XX Sequence 230 AA;
SQ
Query Match 100.0%; Score 342; DB 2; Length 230;
Best Local Similarity 100.0%; Pred. No. 7.3e-31;
Matches 68; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 KSTLPKRAKTKKEQRIERILNRRRAAHQSRKKLHLYLPRKCSLLENLNSVN 60
DB 23 KSTLPKRAKTKKEQRIERILNRRRAAHQSRKKLHLYLPRKCSLLENLNSVN 82
QY 61 LEKLADHE 68
DB 83 LEKLADHE 90
RESULT 6
ADT87049
ID ADT87049 standard; protein; 230 AA.
XX
AC ADT87049;
DT 13-JAN-2005 (first entry)
XX
DE Yeast Stress-related protein from gene YFL052W.
XX
KW Yeast; Stress-related protein; SRP; environmental stress; abiotic stress;
KW drought; heat; cold; salt.
XX
OS Saccharomyces cerevisiae.
XX
XX WO2004092398-A2.
XX
XX 28-OCT-2004.
XX
XX 15-APR-2004; 2004WO-US011888.
XX
XX 15-APR-2003; 2003EP-00008080.
PR 02-MAY-2003; 2003EP-00039728.
PR 01-AUG-2003; 2003EP-00016672.
PR 30-SEP-2003; 2003EP-00022225.
XX
XX (BADI) BASF PLANT SCI GMBH.
XX
XX Puzio P, Chardonnens A, Shirley A, Wang X, Sarria-Millan R;
PI Mckersie B, Chen R;
XX
XX WPI; 2004-766883/75.
DR N-PSDB; ADT87048.
XX
XX New isolated nucleic acid molecule comprises a sequence encoding Stress-
PT Related Protein (SRP), useful for producing transformed plants with
PT altered metabolic activity resulting in increased tolerance or resistance
PT to environmental stress.
XX
XX Claim 15; SEQ ID NO 27; 911pp; English.
PS
XX The invention relates an isolated nucleic acid molecule comprises a
CC nucleic acid molecule encoding a Stress-Related Protein (SRP), from Yeast

CC or E. coli and their homologues from Rice, Soybean and Rape. Also
CC included are a transformed plant cell with altered metabolic activity
CC compared to a corresponding non-transformed wild type plant cell (where
CC the metabolic activity is altered by transformation with a SRP coding
CC nucleic acid and results in increased tolerance and/or resistance to an
CC environmental stress as compared to a corresponding non-transformed wild
CC type plant cell), a transgenic plant generated from the plant cell above
CC (and which is a monocot or dicot plant, or a gymnosperm plant), a seed
CC produced by a transgenic plant above (where the seed is genetically
CC homozygous for a transgene conferring altered metabolic activity
CC resulting in an increased tolerance to environmental stress as compared
CC to a corresponding non-transformed wild type plant), a nucleic acid
CC construct which confers the expression of the nucleic acid molecule above
CC (comprising one or more regulatory elements, where expression of the SRP
CC coding nucleic acid in a host cell results in altered metabolic activity
CC resulting in increased tolerance to environmental stress as compared to a
CC corresponding non-transformed wild type host cell), a vector comprising
CC the nucleic acid molecule above or the nucleic acid construct, a host
CC cell which has been transformed stably or transiently with the vector (or
CC the nucleic acid molecules above, or the nucleic acid construct), an
CC isolated Stress Related Protein (SRP) selected from the amino acid
CC sequences fully given in the specification and/or its homologues, a
CC method of producing a transgenic plant with altered metabolic activity
CC compared to a corresponding non-transformed wild type plant cell,
CC modifying stress tolerance of a plant, detecting environmental stress in
CC plant cells or plants, screening plant cells or plants for increased
CC tolerance and/or resistance to environmental stress, breeding plant cells
CC or plants towards increased tolerance and/or resistance to environmental
CC stress, increasing tolerance of a plant to at least one abiotic stress,
CC a plant transformed with the nucleic acids above and a seed of the plant.
CC The altered metabolic activity and/or a SRP encoding nucleic acids or its
CC homologues are useful as markers for selection of plants or plant cells
CC with increased tolerance to environmental stress, or for detection of
CC stress in plants or plant cells. The nucleic acids are useful for
CC producing transformed plants with altered metabolic activity resulting in
CC increased tolerance and/or resistance to an environmental stress
CC (drought, heat, cold and salt) as compared to a corresponding non-
CC transformed wild-type plant cell. The present sequence is a yeast SRP of
CC the invention.
XX
XX Sequence 230 AA;
SQ

Query Match 100.0%; Score 342; DB 8; Length 230;
Best Local Similarity 100.0%; Pred. No. 7.3e-31;
Matches 68; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 KSTLPKRAKTKKEQRIERILNRRRAAHQSRKKLHLYLPRKCSLLENLNSVN 60
DB 23 KSTLPKRAKTKKEQRIERILNRRRAAHQSRKKLHLYLPRKCSLLENLNSVN 82
QY 61 LEKLADHE 68
DB 83 LEKLADHE 90
RESULT 7
AAW53807
ID AAW53807 standard; protein; 238 AA.
XX
AC AAW53807;
XX
XX 06-JUL-1998 (first entry)
DT
XX
DE Transcription control protein ERN4.
XX
XX Transcription control protein; ERN4; chaperone protein; protein folding;
KW transcription enzyme.
XX
XX Saccharomyces cerevisiae.
OS
XX JPI0084970-A.
PN
XX 07-APR-1998.
PD

```

XX PF 31-OCT-1996; 96JP-00307085.
XX XX
XX PR 23-JUL-1996; 96JP-00213180.
XX XX
XX PA (HSPK-) HSP KENKYUSHO KK.
XX XX
XX DR WPI; 1998-264856/24.
XX DR N-PSDB; AAV20067.
XX XX
XX PT DNA encoding a yeast transcriptional control protein - useful in the
XX PT recombinant production of high yield of protein of interest.
XX PS Claim 2; Page 15; 21pp; Japanese.
XX XX
XX CC This sequence is the yeast transcriptional control protein, designated
XX CC ERN4, of the invention. An alternatively spliced version of the protein
XX CC is shown in AAW52805. Yeast cells transformed with a vector containing
XX CC the DNA can be used for the recombinant production of a protein of
XX CC interest at a high yield. The DNA controls a chaperone protein which in
XX CC turn is responsible for the correct folding of an enzyme involved in
XX CC transcription, hence the cells can be used to control the transcription
XX CC of a recombinant protein of interest
XX SQ Sequence 238 AA;

Query Match 100.0%; Score 342; DB 2; Length 238;
Best Local Similarity 100.0%; Pred. No. 7.6e-31;
Matches 68; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KSTLPKRKAKTEKEQRRIERILNRRRAAHQSRKKRLHLQYLERKCSLLENLNSVN 60
Db |||||
QY 61 LEKLADHE 68
Db |||||
QY 83 LEKLADHE 90
Db |||||

RESULT 8
AAB82975
ID AAB82975 standard; protein; 450 AA.
XX
XX AC AAB82975;
XX
XX DT 11-SEP-2003 (revised)
XX DT 21-DEC-2001 (first entry)
XX
XX DE Trichoderma reesei HAC1, involved in unfolded protein response.
XX KW HAC1; transcription factor; unfolded protein response; protein secretion.
XX OS Hypocrea jecorina.
XX
XX FH Key Location/Qualifiers
XX FT Domain 84..147
XX FT /label= DNA binding domain
XX
XX PN WO200172783-A2.
XX PD 04-OCT-2001.
XX
XX PF 23-MAR-2001; 2001WO-US009401.
XX XX
XX PR 24-MAR-2000; 2000US-00534692.
XX XX
XX PA (GEMV ) GENENCOR INT INC.
XX
XX PI Penttila ME, Ward M, Wang H, Valkonen MJ, Saloheimo MLA;
XX WPI; 2001-626252/72.
XX DR N-PSDB; AAW26931.
XX

```

```

PT Increasing secretion of heterologous proteins e.g. lipase and cellulase
PT in eukaryotic cells useful in industry to increase production and
PT facilitate purification, by inducing an elevated unfolded protein
PT response.
XX
XX PS Claim 54; Fig 7A-B; 89pp; English.
XX
XX CC The present sequence is that of the HAC1 protein of Trichoderma reesei,
XX CC as deduced from the newly isolated HAC1 gene (see AAW26931). HAC1 protein
XX CC is a transcription factor involved in the unfolded protein response
XX CC (UPR). The invention provides methods for increasing the secretion of a
XX CC heterologous protein in a cell by inducing an elevated UPR. This can be
XX CC achieved by modulating the activity of HAC1, PTC2 or IRE1 in the cell,
XX CC e.g. by gene overexpression. The cell from which the protein is secreted
XX CC can be any cell having an UPR, such as mammalian cells, insect cells,
XX CC yeast and filamentous fungi. The protein of interest can be any secreted
XX CC protein such as a therapeutic protein or an industrial enzyme, e.g.
XX CC lipase, cellulase, endoglucanase-H, protease, carbohydratase, reductase,
XX CC oxidase, isomerase, transferase, kinase, phosphatase, alpha-amylase,
XX CC glucosylase, lignocellulose hemicellulase, pectinase and ligninase
XX CC (claimed). (Updated on 11-SEP-2003 to standardise OS field)
XX SQ Sequence 450 AA;

Query Match 57.3%; Score 196; DB 4; Length 450;
Best Local Similarity 61.9%; Pred. No. 7.5e-14;
Matches 39; Conservative 13; Mismatches 9; Indels 2; Gaps 1;

QY 1 KSTLPKRKAKTEKEQRRIERILNRRRAAHQSRKKRLHLQYLERKCSLLENLNSVN 60
Db |||||
QY 102 KTNLPKRKAKTEKEQRRIERILNRRRAAHQSRKKRLHLQYLERKCSLLENLNSVN 159
Db |||||
QY 61 LEK 63
Db ::|
QY 160 VQK 162
Db ::|

RESULT 9
AAE15371
ID AAE15371 standard; protein; 451 AA.
XX
XX AC AAE15371;
XX
XX DT 29-AUG-2003 (revised)
XX DT 07-MAR-2002 (first entry)
XX
XX DE Trichoderma reesei HAC1 protein.
XX
XX KW Heterologous protein secretion; unfolded protein response; UPR; lipase;
XX KW cellulase; carbohydrase; industry; purification; HAC1 protein.
XX
XX OS Hypocrea jecorina.
XX
XX FH Key Location/Qualifiers
XX FT Binding-site 84..147
XX FT /label= DNA-binding_domain
XX
XX PN US2001034045-A1.
XX PD 25-OCT-2001.
XX
XX PF 23-MAR-2001; 2001US-00816277.
XX XX
XX PR 24-MAR-2000; 2000US-00534692.
XX XX
XX PA (GEMV ) GENENCOR INT INC.
XX
XX PI Penttila ME, Ward M, Wang H, Valkonen MJ, Saloheimo MLA;
XX WPI; 2002-033728/04.
XX DR N-PSDB; AAD24595.
XX
XX PT Increasing secretion of heterologous proteins e.g. lipase and cellulase

```


ID ADB31925 standard; protein; 409 AA.
XX AC ADB31925;
XX DT 04-DEC-2003 (first entry)
XX DE Plant (A. thaliana) transcription factor polypeptide #79.
XX KW Plant; transcription factor; transgenic plant; transgenic; plant trait;
XX KW modified trait.
XX OS Arabidopsis thaliana.
XX PN US2003101481-A1.
XX PD 29-MAY-2003.
XX PF 15-NOV-2002; 2002US-00295403.
XX PR 22-SEP-1998; 98US-0101349P.
XX PR 06-OCT-1998; 98US-0103312P.
XX PR 17-NOV-1998; 98US-0108734P.
XX PR 22-DEC-1998; 98US-0113409P.
XX PR 13-SEP-1999; 99US-00394519.
XX PA (ZHAN/) ZHANG J.
XX PA (FROM/) FROMM M.
XX PA (HEAR/) HEARD J.
XX PA (RIEC/) RIECHMANN J L.
XX PA (ADAM/) ADAM L.
XX PA (BROU/) BROUN P.
XX PA (PINE/) PINEDA O.
XX PA (REUB/) REUBER L.
XX PA (KEDD/) KEDDIE J.
XX PA (YUGG/) YU G.
XX PA (JIAN/) JIANG C.
XX PI Zhang J, Fromm M, Heard J, Riechmann JL, Adam L, Broun P;
PI Pineda O, Reuber L, Keddie J, Yu G, Jiang C;
XX WPI; 2003-597572/56.
XX N-PSDB; ADB31924.
XX New isolated polypeptides and polynucleotide sequences, useful for
PT screening a molecule to identify a molecule that modifies plant trait,
PT and for producing plants with modified traits.
XX Claim 9; SEQ ID NO 158; 17pp; English.
XX The present invention relates to the isolation of plant (Arabidopsis
CC thaliana) transcription factor polypeptide and polynucleotide sequences.
CC Also disclosed are: an expression vector comprising the isolated
CC polynucleotide, a host cell comprising the expression vector, a
CC transgenic plant comprising the isolated polynucleotide, a transgenic
CC plant ectopically expressing the isolated polynucleotide or polypeptide,
CC a method for screening a molecule to identify a molecule that modifies a
CC plant trait by placing the molecule in contact with the plant, and
CC monitoring the effect of the molecule on the expressing or activity of
CC the polypeptide or polynucleotide, and producing a transgenic plant
CC having a modified trait by ectopically expressing the isolated
CC polypeptide and selecting a plant with the modified trait. The
CC polypeptides, polynucleotides and methods are useful for screening a
CC molecule to identify a molecule that modifies plant trait, and for
CC producing plants with modified traits. The present sequence represents a
CC plant transcription factor polypeptide of the invention. Note: The
CC sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC USPTO web site at seqdata.uspto.gov.
XX Sequence 409 AA;
SQ

Query Match 52.2%; Score 178.5; DB 7; Length 409;
Best Local Similarity 56.5%; Pred. NO. 6.8e-12;

Matches 39; Conservative 14; Mismatches 13; Indels 3; Gaps 2;
Qy 1 KSTLPPPKRAKTKKEQRRIRILNRRAAHOSSREKKRLHLQYLERKCSLLENLLSVN 60
Db 84 KTNLPPPKRAKTKQDEKQRRVERVLNRRAAQSSRRKRQGEVALEVEKRAIER--KNMD 141
Qy 61 LE-KLADHE 68
Db 142 LEMRLADME 150
RESULT 14
ADO02271
ID ADO02271 standard; protein; 409 AA.
XX AC ADO02271;
XX DT 01-JUL-2004 (first entry)
XX DE Thalecress transcription factor protein #342.
XX KW Thalecress; transcription factor; plant; transgenic; abiotic stress;
KW cold tolerance; heat tolerance; drought; osmotic stress;
KW phosphate limitation; potassium limitation; nitrogen limitation;
KW hormone sensitivity; disease resistance; sugar sensing; seed germination;
KW flowering; inflorescence architectural change;
KW meristem cell differentiation; phyllotaxy; apical dominance;
KW trichome development; seed development; premature senescence;
KW delayed senescence; lethality; necrosis; plant size; leaf morphology;
KW seed morphology; secondary metabolism; light response; shade avoidance.
XX Arabidopsis thaliana.
XX OS Arabidopsis thaliana.
XX PN US2004045049-A1.
XX PD 04-MAR-2004.
XX PF 10-APR-2003; 2003US-00412699.
XX PR 13-SEP-1999; 99US-00394519.
XX PR 21-JAN-2000; 2000US-00489376.
XX PR 17-FEB-2000; 2000US-00506720.
XX PR 22-MAR-2000; 2000US-00532591.
XX PR 22-MAR-2000; 2000US-00533029.
XX PR 22-MAR-2000; 2000US-00533030.
XX PR 22-MAR-2000; 2000US-00533392.
XX PR 22-MAR-2000; 2000US-00533648.
XX PR 06-APR-2000; 2000WO-US009448.
XX PR 16-NOV-2000; 2000US-00713994.
XX PR 17-MAR-2001; 2001US-00819142.
XX PR 17-APR-2001; 2001US-00837444.
XX PR 30-JAN-2002; 2002US-00958131.
XX PR 14-JUN-2002; 2002US-00171468.
XX PR 09-AUG-2002; 2002US-00225066.
XX PR 09-AUG-2002; 2002US-00225067.
XX PR 09-AUG-2002; 2002US-00225068.
XX PR 17-DEC-2002; 2002US-0434166P.
XX PR 25-FEB-2003; 2003US-00374780.
XX (ZHAN/) ZHANG J.
XX PA (FROM/) FROMM M E.
XX PA (HEAR/) HEARD J E.
XX PA (RIEC/) RIECHMANN J L.
XX PA (ADAM/) ADAM L J.
XX PA (BROU/) BROUN P E.
XX PA (PINE/) PINEDA O.
XX PA (REUB/) REUBER T L.
XX PA (KEDD/) KEDDIE J S.
XX PA (YUGG/) YU G.
XX PA (JIAN/) JIANG C.
XX PA (SAMA/) SAMAHA R S.
XX PA (PILG/) PILGRIM M L.
XX PA (CREE/) CREELMAN R A.

	Matches	38;	Conservative	12;	Mismatches	17;	Indels	1;	Gaps	1;
Qy	1	KSTLPPRKRAKTEKEQRIERILNRRAAHQSREKRLHLQYLERKCSLLENLLNSVN 60								
Db	64	KTNLPPRKRAKTEDEKEQRIERVLNRFAAAQTSREKRLKLEMEKLENEKIQMEQ-QNQFL 122								
Qy	61	LEKLADHE 68								
Db	123	LQRLSQME 130								

Search completed: November 23, 2005, 03:18:41
Job time : 27.5576 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 23, 2005, 03:18:52 ; Search time 4.42281 Seconds
(without alignments)
239.171 Million cell updates/sec

Title: US-10-663-450-4

Perfect score: 1766

Sequence: 1 MKSADRFSPVKMEDAFANSP.....PSHGASTSRCGQIAAGSA 349

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 17545 seqs, 3030971 residues

Total number of hits satisfying chosen parameters: 17545

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA.New:

- 1: /cgn2_6/ptodata/2/pubpa/US10_NEW_PUB_PEP.*
- 2: /cgn2_6/ptodata/2/pubpa/US06_NEW_PUB_PEP.*
- 3: /cgn2_6/ptodata/2/pubpa/US07_NEW_PUB_PEP.*
- 4: /cgn2_6/ptodata/2/pubpa/US08_NEW_PUB_PEP.*
- 5: /cgn2_6/ptodata/2/pubpa/US09_NEW_PUB_PEP.*
- 6: /cgn2_6/ptodata/2/pubpa/PCT_NEW_PUB_PEP.*
- 7: /cgn2_6/ptodata/2/pubpa/US11_NEW_PUB_PEP.*
- 8: /cgn2_6/ptodata/2/pubpa/US60_NEW_PUB_PEP.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	192	10.9	376	1	US-10-971-994-2
2	131.5	7.4	431	1	US-10-821-234-1285
3	121.5	6.9	605	1	US-10-131-826A-160
4	121	6.9	588	7	US-11-022-562-210
5	118.5	6.7	1618	1	US-10-984-645-2
6	113	6.4	1138	1	US-10-509-422-4
7	104.5	5.9	1168	1	US-10-509-422-2
8	102.5	5.8	1616	1	US-10-821-234-1497
9	102	5.8	298	1	US-10-821-234-1240
10	101	5.7	4384	1	US-10-821-234-1120
11	100.5	5.7	2432	1	US-10-821-234-899
12	98	5.5	431	7	US-11-169-013-2
13	98	5.5	454	7	US-11-169-013-1
14	97.5	5.5	216	1	US-10-821-234-925
15	97	5.5	865	7	US-11-109-156-4
16	97	5.5	1122	1	US-10-821-234-1657
17	96	5.4	401	1	US-10-821-234-881
18	95	5.4	1410	1	US-10-821-234-1050
19	94	5.3	676	7	US-11-135-855-28
20	94	5.3	717	7	US-11-135-855-29
21	93.5	5.3	413	1	US-10-967-648A-8
22	92	5.2	777	1	US-10-821-234-1658
23	91.5	5.2	135	1	US-10-821-234-1037
24	91.5	5.2	1377	1	US-10-821-234-1070
25	91	5.2	1142	7	US-11-044-051-73

ALIGNMENTS

RESULT 1

US-10-971-994-2
; Sequence 2, Application US/10971994
; Publication No. US20050250182A1
; GENERAL INFORMATION:
; APPLICANT: University of Michigan et al.
; APPLICANT: Kaufman, Randal
; APPLICANT: Kyung, Lee
; APPLICANT: Mori, Kazutos
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, AND METHODS FOR
; MODULATING THE UNFOLDED PROTEIN RESPONSE
; FILE REFERENCE: UMV-2246
; CURRENT APPLICATION NUMBER: US/10/971,994
; CURRENT FILING DATE: 2004-10-21
; PRIOR APPLICATION NUMBER: PCT/2003/012640
; PRIOR FILING DATE: 2003-04-22
; PRIOR APPLICATION NUMBER: 60/375,098
; PRIOR FILING DATE: 2002-04-22
; PRIOR APPLICATION NUMBER: 60/374,880
; PRIOR FILING DATE: 2002-04-23
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2

LENGTH: 376

TYPE: PRT

ORGANISM: Homo sapiens

US-10-971-994-2

Query Match 10.9%; Score 192; DB 1; Length 376;

Best Local Similarity 26.0%; Pred. No. 2.6e-08;

Matches 93; Conservative 52; Mismatches 127; Indels 86; Gaps 17;

QY	33	SPADTSLRKNVVAQPKPEKKPAKKRSGQELP--VP-----KTNLP-PRKRA	79
DB	9	NPADGTPKVLISGGQPASAGAPA-----GQALPLMVPAQRGASPEASGGLPQARKQ	62
QY	80	K-TEDEKQRRIVRLNRNRAAQTGRKRLEMEKLESEKIDMEQOQFL-----QR	131
DB	63	RUTLSPEKALRRKLKNRVAQTADRKAKRMESELEQQVVDLEENQKLLLENQKLLREK	122
QY	132	LAQMEANNRLSQO-----VAQLSAEVGSG--RHSTPTSSSPASVSPITLPTLFLQEGD	183
DB	123	THGLVVENQELRQLGMDALVAEEAEAKGNEVRPVAGSAEAGAGPVVTPP-----E	176
QY	184	EYPLDR-----IPPTPSVTDVSPYTLKPSLSLAESPDLTQHPAVSVGGLE	227
DB	177	HLPMDGGIDSSDSSEDLILGILDNPVNFKPCSPFASLEELFEV--YP-----E	227

Db 376 TQVICKDMSNLAKGNYIILNWTENIDCEVFRQHRGPOLLALVEVLPRHGS 426

RESULT 4

US-11-022-562-210
; Sequence 210, Application US/11022562
; Publication No. US20050249742A1
; GENERAL INFORMATION:
; APPLICANT: Ruprecht, Ruth M.
; APPLICANT: Shisong, Jiang
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR MODULATING
; TITLE OF INVENTION: A CYTOTOXIC T LYMPHOCYTE IMMUNE RESPONSE
; FILE REFERENCE: DEN-043CN
; CURRENT APPLICATION NUMBER: US/11/022,562
; CURRENT FILING DATE: 2004-12-22
; PRIOR APPLICATION NUMBER: PCT/US03/20322
; PRIOR FILING DATE: 2003-06-27
; PRIOR APPLICATION NUMBER: 60/392718
; PRIOR FILING DATE: 2002-06-27
; NUMBER OF SEQ ID NOS: 340
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 210
; LENGTH: 588
; TYPE: PRT
; ORGANISM: Ebola virus
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: 120
; OTHER INFORMATION: Xaa = Any Amino Acid
US-11-022-562-210

Query Match 6.9%; Score 121; DB 7; Length 588;
Best Local Similarity 21.6%; Pred. No. 0.014;
Matches 61; Conservative 38; Mismatches 103; Indels 80; Gaps 9;

Qy 91 ERLVRNRAAQTSSRERKLEMEKLESEKIDMEQNOFLQRLAQMEANRLSQOVAQLS 150
Db 237 EQVQQLREATAEKQQAESRELDHGLDQEKILMNFHQ---KNEISFQQTNAM 293
Qy 151 AEVGRSRHSTPTSSPASVSFTLTPTLTKQEGDEVPLDRIPPTPSVDTYSPTLKPSSLA 210
Db 294 VTLRKERLAKLTAITAASLP-----KTSGHVDDDDIPFPPIINDDONPGHQDDPT 346
Qy 211 ESPDLT-----QHPAVSVGLGEGDESALTLPDL----- 238
Db 347 DSQDTTIPVVVDPDDGSYGEYSYSENGMNAFDD-LVLPDLDEDDDEDTKPVNRSYTKGG 405
Qy 239 -----GASIKHEPT-----HDLTAPLSDDFRRLFNGLSDSSLSLL 276
Db 406 QQKNSQKQGTGRTQRTQNVPGPHRTIHASAPLTNDNRNPSGTSRMLTPIN 465
Qy 277 EDGFAFDVLDSGD--LSAFPFDMSVDPDTEPVTLEDLQTNQ 316
Db 466 EEA--DPLDADDETSLSL-----PLSSDDEEQDRG 494

RESULT 5

US-10-984-645-2
; Sequence 2, Application US/10984645
; Publication No. US20050244386A1
; GENERAL INFORMATION:
; APPLICANT: Habener, Joel
; APPLICANT: Zulewski, Hendrik
; APPLICANT: Abraham, Elizabeth
; APPLICANT: Vallejo, Mario
; TITLE OF INVENTION: METHOD OF TRANSPLANTING IN A MAMMAL AND TREATING DIABETES MELLITU
; TITLE OF INVENTION: BY ADMINISTERING A PSEUDO-ISLET LIKE AGGREGATE DIFFERENTIATED FR
; FILE REFERENCE: 3284/1223
; CURRENT APPLICATION NUMBER: US/10/984,645
; CURRENT FILING DATE: 2004-11-09
; - PRIOR APPLICATION NUMBER: US 09/731,255
; PRIOR FILING DATE: 2000-12-06

; PRIOR APPLICATION NUMBER: US 60/169,082
; PRIOR FILING DATE: 1999-12-06
; PRIOR APPLICATION NUMBER: US 60/215,109
; PRIOR FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: US 60/239,880
; PRIOR FILING DATE: 2000-10-06
; NUMBER OF SEQ ID NOS: 55
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 2
; LENGTH: 1618
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-984-645-2

Query Match 6.7%; Score 118.5; DB 1; Length 1618;
Best Local Similarity 21.9%; Pred. No. 0.085;
Matches 85; Conservative 69; Mismatches 152; Indels 83; Gaps 20;

Qy 8 SPVKMEDAFANSTPTSPSLEVPVLTVPADTSLRTKNVVAQTKPEKKPAKRRKSWGOEL- 66
Db 679 SP-EVGDEALRPLTKENOEPLRSLED-----ENKEAFRSLEKENOEPLKTLSEEDOSIV 732
Qy 67 -PVPKTNLPKRKAKTEDEKEQRRIERVLNRNAAA-----QTSRERKRLEMEKLES--E 117
Db 733 RPLETENHKSLSRLEQDOQETLTLKETQORRSLSGEQDQMTLRPEKVDLSPKSLDQ 792
Qy 118 KI--DMEQNOQLQRLAQMEANRLSQOVAQLSAEVRGS-----RHSTPTSSPASVS 170
Db 793 ETARPLENENQFLKSLKEESVE-----AVKSLETEILLESLSKAGQENLETILKSPETA 846
Qy 171 PTLTP--TLFKQSGDE-----VPLDRIPFPPTP-----SVTDYSPTLKPSL 209
Db 847 PLWTPBEINKSGNNESSRKNSTTVCVCGSEPRD---IOTPGGESGIIISGSMPEGEF 903
Qy 210 AASPDLTQ-----HPAVSVGLGEGDESALTLPDLGASI-----KHEPTHDLTAPLSD 256
Db 904 ELSRGVDKESQRLNEEENLGRGEQFSLRSLEEGQELPQSADVQVRWEDTVEKQQLAQ 963
Qy 257 DFRRLFNGLSDSSLSLE--DGFA--FDVLDSGLSA-----PFSDMSVDPDTEPVTL 308
Db 964 ESPPGMAGVNDKAEALNLRQDQGTGKEVVEQGLNATEEVRWFPPEG-----HPENP 1017
Qy 309 EDLEQTNGLSDASCK--AASLPQSHGAS 335
Db 1018 EPKEQ-RGLVEGASVKGAGLQDPPEGQS 1045

RESULT 6

US-10-509-422-4
; Sequence 4, Application US/10509422
; Publication No. US20050244825A1
; GENERAL INFORMATION:
; APPLICANT: Liou, Simon
; TITLE OF INVENTION: Human BMP2 Inducible Kinases
; FILE REFERENCE: 004974.01015
; CURRENT APPLICATION NUMBER: US/10/509,422
; CURRENT FILING DATE: 2004-09-24
; PRIOR APPLICATION NUMBER: PCT/EP03/080825
; PRIOR FILING DATE: 2003-03-20
; PRIOR APPLICATION NUMBER: US 60/367,512
; PRIOR FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: US 60/406,936
; PRIOR FILING DATE: 2002-08-30
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 1138
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-509-422-4

Query Match 6.4%; Score 113; DB 1; Length 1138;
Best Local Similarity 20.8%; Pred. No. 0.14;

	Matches	75; Conservative	63; Mismatches	163; Indels	60; Gaps	14;
Qy	17	ANSPPTPSLEVPVLTVSPADTSURTKNVVAQTKEPKP-AKRKSGOELPV---	BKTN	72		
Dd	373	ANSTAATS---SVLTIQSSTAPVK---VPAGFEFSNHKPGALRPGNGSEVLMVYGGPPQQ	426			
Qy	73	LPRKRAKTEDEKEORRIERVLRNRAAAQTSRRERKLEMEKLSEKIDMQEQNQFLIQLRL	132			
Dd	427	PPOQHVRLOLQOGDMRWLOLHLHRHPHHHQOQQOQQOQQOQQOQQOQQOQQOQQOQQO	486			
Qy	133	AQMAEANNRLSQOVAQLSAEVRGSRHSTPTSSPASVSPTLT-----TLFKQ	180			
Dd	487	YLQYQHAMHQHILQOQFLM----HSVYQOPPPASQYPAMMQYQAFTLQOQMARHQ	542			
Qy	181	EGDEVPLDRIPFP---TPSVTDYSPTLKPSLSAESPDLTQHPAVSVGGLEGDESALTFD	237			
Dd	543	PAQQVSPEYLTSPQFSPALVSAYSSL-PAQVGITVDVSYCANRSV---AEKEAVANFT	597			
Qy	238	LGSASIKHEPTHDLTAPLSDDDFRLLFNGLDSSLESDDSLLEDGFAPFVLDGSDILSA-PFFD	296			
Dd	598	NQKTI SHPPDMGWNPFGEENDFSKL-----TEEELDR--EFDLLRNRLGASTPSP	647			
Qy	297	SMWDFOTEPVTL---EDLEQTNGLSDSASCKAASLOPS-----HGASTSCDG	341			
Dd	648	KTVDLPPAHSRPPEEPFASVPFISSHSGSPEKKTTTSHPNQKSITANLTKNNGSSPLCKD	707			
Qy	342	Q	342			
Dd	708	O	708			

```

RESULT 7
US-10-509-422-2
; Sequence 2, Application US/10509422
; Publication NO. US20050244825A1
; GENERAL INFORMATION:
; APPLICANT: Liou, Simon
; TITLE OF INVENTION: Human BMP2 Inducible Kinases
; FILE REFERENCE: 004974.01015
; CURRENT APPLICATION NUMBER: US/10/509,422
; CURRENT FILING DATE: 2004-03-24
; PRIOR APPLICATION NUMBER: PCT/EP03/080825
; PRIOR FILING DATE: 2003-03-20
; PRIOR APPLICATION NUMBER: US 60/367,512
; PRIOR FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: US 60/406,936
; PRIOR FILING DATE: 2002-08-30
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 1168
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-509-422-2

```

Query Match	5.9%	Score 104.5	DB 1	Length 1168
Best Local Similarity	20.8%	Pred. No. 0.65		
Matches	64	Conservative	57	Mismatches 129; Indels 57; Gaps 12;
QY	17	ANSPT--TPSLEVPVLTVSADTSLRTKNVVAOTKPEKKP-AKKRKSQGQELPV---PK	70	
Db	376	ANSATTATPS-----VLTIQSSATPVK---VLAPGEFGNHRPKGALRPGNGPEILLGGQPP	428	
QY	71	TNLPPrKRAKTEDEKEORRIERV-LRNRAAAOTSRERKLEMEKLESEKIDMEQO---	126	
Db	429	QQPQQHRVQLQQLGGDWRLQQLHQRHPHQOQOQOQOQOQOQOQOQOQOQOQH	488	
QY	127	-----FLQLRLQWAEENRLSQVAQLSAEVRGSRHSTPTSSPASVSPTLTTLPKQE	181	
Db	489	HHHHHLLQDAYQQQYQHATQQOQOMLQQQFLMHSVYQPFPSAQYPTTMPEYQQQAFQQQ	548	
QY	182	GDEVPLDRIPFPPTPSVTDYSPTLKPSSLAESPDLTQHPA---VSVGGI-----	226	

```

Db      549  -----MLAQHPSQQQASPEYLTSPOEFSPALVSVTSSLPAQVGTIMDSSYSANRQVF 601
Qy      227  ---EGDESALTFLDGLGASIKHEPHTDITAPLSDDDFRRLFNGLSDSSLESDDLSDGFAPD 283
Db      602  FQSVADKEAIAINTNOKNISNPDMGWNPFGE DNFSKL-----TEBELLDLDR--EFD 651
Qy      284  VLDSDGL 290
Db      652  LLRSNRL 658

RESULT 8
US-10-821-234-1497
; Sequence 1497, Application US/10821234
; Publication No. US20050255114A1
; GENERAL INFORMATION:
; APPLICANT: Labat, Ivan
; APPLICANT: Stache-Crain, Birgit
; APPLICANT: Andarmani, Susan
; APPLICANT: Tang, Y. Tom
; TITLE OF INVENTION: Methods for Diagnosis and Treatment of Preeclampsia
; FILE REFERENCES: 821A
; CURRENT APPLICATION NUMBER: US/10/821,234
; CURRENT FILING DATE: 2004-04-07
; PRIOR APPLICATION NUMBER: US 60/462,047
; PRIOR FILING DATE: 2003-04-07
; NUMBER OF SEQ ID NOS: 1704
; SOFTWARE: pt_seq_genes Version 1.0
; SEQ ID NO 1497
; LENGTH: 1616
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-821-234-1497

```

Query Match	5.8%;	Score 102.5;	DB 1;	Length 1616;
Best Local Similarity	22.1%;	Pred. No. 1.4;		
Matches	80;	Conservative 40;	Mismatches 113;	Indels 129; Gaps 18;
Qy	15	AFANSPPTTSLEVPVLTVSPADTSLRTKNNVVAQTKPE-----	-----EK	53
Db	123	ADANSPPKP-----LSKPRTPRSKS--DGEAKPEPSPRITRKSTRTQTTITSHFAK	173	
Qy	54	KPAKRRKSGQELPVPKTNLPPIPKRAKTEDEKQORRI---ERVLR-----NRAAAQTS	103	
Db	174	GPAKRRKPQ--EESERAKSDESIKEEDKDQDEK--RRVTSRERVARPLPABEPERAKSGT-	229	
Qy	104	RERKRLEMEKLESEKIDMEQONQFLLRQAQM-----EAENNRLSQQ	145	
Db	230	RTKEBERDEKEKRLRSQKTEPTPQKLEEDPREARAGVQADEDEGDEKDEKKHRSQ	289	
Qy	146	VAOLSAEVRGSRHSTPTSSPASVSTLTPTLFKQEGDEVLPDRIPPTPSVTIDYSTLK	205	
Db	290	PKDLAAKRR-----PEEKEPEKVP-----QISDEKDEDEKEEKRKKTTPKEPTEK	335	
Qy	206	PSLSA-----ESPDLT--QHPAVSVGGLEGDE-----SALTLED	237	
Db	336	KMARAKTVNMSKTHPPKCIQCGQYLDPPDLUKYGOHPDVA-----DEPQMLTNEKLSIFD	390	
Qy	238	LG-----ASIKHEPTHDLT-----APLSDDDFRRLFNQDSS--LESDDSSILLE	277	
Db	391	ANESGFESYEALPQHKLTCFVSVCYKHCHLCPIDTGLIEKNIUEFFSGSAKPIYDDPSLE	450	
Qy	278	DG 279		
Db	451	GG 452		

RESULT 9
US-10-821-234-1240
; Sequence 1240, Application US/10821234
; Publication No. US2005025114A1
; GENERAL INFORMATION:
; APPLICANT: Labat, Ivan


```
; APPLICANT: Stache-Crain, Birgit
; APPLICANT: Andarmani, Susan
; APPLICANT: Tang, Y. Tom
; TITLE OF INVENTION: Methods for Diagnosis and Treatment of Preclampsia
; FILE REFERENCE: 821A
; CURRENT APPLICATION NUMBER: US/10/821,234
; CURRENT FILING DATE: 2004-04-07
; PRIOR APPLICATION NUMBER: US 60/462,047
; PRIOR FILING DATE: 2003-04-07
; NUMBER OF SEQ ID NOS: 1704
; SOFTWARE: pt_seq_genes Version 1.0
; SEQ ID NO 1240
; LENGTH: 298
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-821-234-1240

Query Match          5.8%; Score 102; DB 1; Length 298;
Best Local Similarity 27.0%; Pred. No. 0.16;
Matches 41; Conservative 18; Mismatches 49; Indels 44; Gaps 6;

QY 15 AFANSPPTPSLEVPLTVSPADTSRTKNVVAQTPEEKPAKRSWGQELPVPKTNLP 74
Db 179 AAAGOPTPTSPPEPRSS-SPROT-----PAPGPAREKSAGKRG-----PDGSGP 221
QY 75 PKRAKTEDEKEQRRIERVLNRNAAQTSRERKRLEMEKLESEKIDMEQOQFLQRLAQ 134
Db 222 -----EYQRR-----ERNNIAVRKSRDKAK-----RRNQEMQOKLVE 254
QY 135 MEAENNRLSQQVAQLSAEVRGSRHSTPTSSSP 166
Db 255 LSAENEKHLHQRVEQLTRDLAGLQRFQFKQLPSP 286

RESULT 10
US-10-821-234-1120
; Sequence 1120, Application US/10821234
; Publication No. US20050255114A1
; GENERAL INFORMATION:
; APPLICANT: Labat, Ivan
; APPLICANT: Stache-Crain, Birgit
; APPLICANT: Andarmani, Susan
; APPLICANT: Tang, Y. Tom
; TITLE OF INVENTION: Methods for Diagnosis and Treatment of Preclampsia
; FILE REFERENCE: 821A
; CURRENT APPLICATION NUMBER: US/10/821,234
; CURRENT FILING DATE: 2004-04-07
; PRIOR APPLICATION NUMBER: US 60/462,047
; PRIOR FILING DATE: 2003-04-07
; NUMBER OF SEQ ID NOS: 1704
; SOFTWARE: pt_seq_genes Version 1.0
; SEQ ID NO 1120
; LENGTH: 4384
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-821-234-1120

Query Match          5.7%; Score 101; DB 1; Length 4384;
Best Local Similarity 21.9%; Pred. No. 7.4;
Matches 73; Conservative 54; Mismatches 139; Indels 68; Gaps 16;

QY 5 DRFSVKMEDAFANSPPTPSLEVPL-TVSPADTSRTKNVVAQTPEEKPAKRSWG 63
Db 3158 DTLEQVSLDSGSGKSPLP--ETPSSESVSEFTSKTPDSLAIYIP-----G 3202
QY 64 QELPVKTN--LPPRKRAKTEDEK---EORRIERVLNRNAAQTSRERKRLEM----- 111
Db 3203 KPSPIPEVSESEEEQAKSTLSKQTTVEETFAVEREMENDVSKNSQNRPNRVAYIEFP 3262
QY 112 --EKLESEKIDMEQOQFLQRLAQMEAENNRLSQQVAQLSAEVRGSRHSTPTSSSP-AS 168
Db 3263 PPPPLDADQIESDKKHLYLPEKVDWIEVNLQDEHDKYLAEPV--IRVQPPSPVPPGAD 3320

; APPLICANT: Stache-Crain, Birgit
; APPLICANT: Andarmani, Susan
; APPLICANT: Tang, Y. Tom
; TITLE OF INVENTION: Methods for Diagnosis and Treatment of Preclampsia
; FILE REFERENCE: 821A
; CURRENT APPLICATION NUMBER: US/10/821,234
; CURRENT FILING DATE: 2004-04-07
; PRIOR APPLICATION NUMBER: US 60/462,047
; PRIOR FILING DATE: 2003-04-07
; NUMBER OF SEQ ID NOS: 1704
; SOFTWARE: pt_seq_genes Version 1.0
; SEQ ID NO 1240
; LENGTH: 298
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-821-234-899
; Sequence 899, Application US/10821234
; Publication No. US20050255114A1
; GENERAL INFORMATION:
; APPLICANT: Labat, Ivan
; APPLICANT: Stache-Crain, Birgit
; APPLICANT: Andarmani, Susan
; APPLICANT: Tang, Y. Tom
; TITLE OF INVENTION: Methods for Diagnosis and Treatment of Preclampsia
; FILE REFERENCE: 821A
; CURRENT APPLICATION NUMBER: US/10/821,234
; CURRENT FILING DATE: 2004-04-07
; PRIOR APPLICATION NUMBER: US 60/462,047
; PRIOR FILING DATE: 2003-04-07
; NUMBER OF SEQ ID NOS: 1704
; SOFTWARE: pt_seq_genes Version 1.0
; SEQ ID NO 899
; LENGTH: 2432
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-821-234-899

Query Match          5.7%; Score 100.5; DB 1; Length 2432;
Best Local Similarity 21.9%; Pred. No. 3.6;
Matches 86; Conservative 53; Mismatches 147; Indels 107; Gaps 16;

QY 14 DAFANSPPTPSLEV-----PVLTVSPADTSRTKN-----VVAQTKPEEKPAK 57
Db 54 DAASASRLPNEEIVQKIEEVLG-VLDTLYRYKPDLEKSGRSKRCVSVQTDPTDEIPTK 112
QY 58 KRKSWGQELPVKTNLPPRKRAKTEDEKEQRRIERVLNRNAAQTSRERKRLEMEK- 113
Db 113 KSKK-----HKHKNNKKKKKKKKKKYKRPQPESESKT---KSHDDGNIDLESDSL 162
QY 114 -----LESEKIDMEQOQFLQRL--AQMEAENNR 141
Db 163 KFDSEPSAVALLELPTRAFGPSETNESPAVVLEPPVVGMEVSEPHILETLKPAKTAELSV 222
QY 142 LSQOVAQLSAEVRGSRHSTPT-----SSSPASVSPTLTPTLTKOEGDEVPLDRIPFP 196
Db 223 VTSVISEQSESVAVMPEPSMTKILDSFAAAPVPTTLVLKSE-----PVVTMS 273
QY 197 VTDYSPKLPSSLAESPD-----LTQHPAVSVGGLGDESALTFLDLGASIKHETHOLT 251
Db 274 VEYQMKSVLKSVESTSPSPSKIMLVEPPVAKV--LEPSETLWVSSETPTVVPPEPSTST 331
QY 252 APLSDDDFRRLFNCGDSSLSLESDGFADFVLDG--DLSAFPFDSDMVDFTDEPTVLED 310
Db 332 M-----DF-----PESSAIE---AURLPEQPVDPVSEIADSSM---TRQELPE 369
QY 311 LEQTNGLSDSASCKAASLQPSHGASTSRCDGQG 343
Db 370 LPKTTALEQESSVASAMELPGPPATSMPELOG 402

RESULT 12
US-11-169-013-2
; Sequence 2, Application US/11169013
; Publication No. US20050244971A1
```

```

; APPLICANT: Andarmani, Susan
; APPLICANT: Tang Y. Tom
; TITLE OF INVENTION: Methods for Diagnosis and Treatment of Preeclampsia
; FILE REFERENCE: 821A
; CURRENT APPLICATION NUMBER: US/10/821,234
; CURRENT FILING DATE: 2004-04-07
; PRIOR APPLICATION NUMBER: US 60/462,047
; PRIOR FILING DATE: 2003-04-07
; NUMBER OF SEQ ID NOS: 1704
; SOFTWARE: pt_seq_genes Version 1.0
; SEQ ID NO 925
; LENGTH: 216
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(216)
; OTHER INFORMATION: Xaa = any amino acid or nothing
US-10-821-234-925

Query Match          5.5%; Score 97.5; DB 1; Length 216;
Best Local Similarity 25.4%; Pred. No. 0.22; Indels 31; Gaps 6;
Matches 45; Conservative 22; Mismatches 79;

QY 47 QTPPEKKPAKRRKSGOELPVKTNLPPIKRAKTEDEKEORRIERVLNRNRAAAQTSSR 106
DB 32 RTTPEQ-----QERCGPHITSKRQRPTREQEKHTQ-----RGAAQQNRRKQ 74

QY 107 KRLMEKLI-ESEKIDMEQQNQFLLQRLAOMEAENRLSQVQAUSAEVGRSHSTPTSSS 165
DB 75 NHGERTFTQSKSDSDKKQ-----QKSSHRTTTQXHKPEQORTA--TTPORQA 123

QY 166 PASVSPFLTPTLPKQEGDEVPLD--RIPFPPTSPVTDYSPLTKPSSLAESPDLTQHPA 220
DB 124 DTRTRSGTSTTKREQTPDEARARRRISKSTRHTTETKATTTTRRDANEHTHTTORPA 180

RESULT 15
US-11-109-156-4
; Sequence 4, Application US/1109156
; Publication No. US20050250144A1
; GENERAL INFORMATION:
; APPLICANT: Toshio Ota
; APPLICANT: Takao Isogai
; APPLICANT: Tetsuo Nishikawa
; APPLICANT: Koji Hayashi
; APPLICANT: Kaoru Otsuka
; APPLICANT: Jun-Ichi Yamamoto
; APPLICANT: Shizuko Ishii
; APPLICANT: Tomoyasu Sugiyama
; APPLICANT: Ai Wakamatsu
; APPLICANT: Keiichi Nagai
; APPLICANT: Tetsuji Otsuki
; APPLICANT: Shin-Ichi Funahashi
; APPLICANT: Chiaki Senoo
; APPLICANT: Jun-Ichi Nezu
; TITLE OF INVENTION: NOVEL GENES ENCODING PROTEIN KINASE/PROTEIN
; TITLE OF INVENTION: PHOSPHATASE
; FILE REFERENCE: 06501-099002
; CURRENT APPLICATION NUMBER: US/11/109,156
; CURRENT FILING DATE: 2005-04-19
; PRIOR APPLICATION NUMBER: US/10/060,065
; PRIOR FILING DATE: 2002-01-29
; PRIOR APPLICATION NUMBER: PCT/JP00/05061
; PRIOR FILING DATE: 2000-07-28
; PRIOR APPLICATION NUMBER: US 60/159,590
; PRIOR FILING DATE: 1999-10-18
; PRIOR APPLICATION NUMBER: US 60/183,322
; PRIOR FILING DATE: 2000-02-17
; PRIOR APPLICATION NUMBER: JP 11-248036
; PRIOR FILING DATE: 1999-07-29
; PRIOR APPLICATION NUMBER: JP 2000-118776
; PRIOR FILING DATE: 2000-01-11

```


THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 23, 2005, 03:12:56 ; Search time 112.993 Seconds
(without alignments)
1290.661 Million cell updates/sec

Title: US-10-663-450-4
Perfect score: 1766
Sequence: 1 MKSADRFSPVKMEDAFANSP.....PSHGASTSRCDGGGIAAGSA 349

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications AA Main:*

- 1: /cgn2_6/prodata/1/pubpaa/US07_PUBCOMB.pgp:*
- 2: /cgn2_6/prodata/1/pubpaa/US08_PUBCOMB.pgp:*
- 3: /cgn2_6/prodata/1/pubpaa/US09_PUBCOMB.pgp:*
- 4: /cgn2_6/prodata/1/pubpaa/US10A_PUBCOMB.pgp:*
- 5: /cgn2_6/prodata/1/pubpaa/US10B_PUBCOMB.pgp:*
- 6: /cgn2_6/prodata/1/pubpaa/US11_PUBCOMB.pgp:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1766	100.0	349	US-09-816-277-4	Sequence 4, Appl1
2	1766	100.0	349	US-09-816-277-6	Sequence 6, Appl1
3	1766	100.0	349	US-10-663-450-4	Sequence 4, Appl1
4	1766	100.0	349	US-10-663-450-6	Sequence 6, Appl1
5	1293.5	73.2	342	US-09-816-277-19	Sequence 19, Appl1
6	1293.5	73.2	342	US-10-663-450-19	Sequence 19, Appl1
7	1293.5	73.2	386	US-09-816-277-16	Sequence 16, Appl1
8	1293.5	73.2	386	US-10-663-450-16	Sequence 16, Appl1
9	434	24.6	451	US-09-816-277-2	Sequence 2, Appl1
10	434	24.6	451	US-09-816-277-5	Sequence 5, Appl1
11	434	24.6	451	US-10-663-450-2	Sequence 2, Appl1
12	434	24.6	451	US-10-663-450-5	Sequence 5, Appl1
13	398.5	22.6	409	US-09-533-029-104	Sequence 104, Appl1
14	398.5	22.6	409	US-10-295-403-158	Sequence 158, Appl1
15	398.5	22.6	409	US-10-412-699B-684	Sequence 684, Appl1
16	318	18.0	174	US-10-767-701-61238	Sequence 61238, A
17	216	12.2	200	US-10-369-493-21867	Sequence 21867, A
18	176.5	10.0	68	US-09-816-277-60	Sequence 60, Appl1
19	176.5	10.0	68	US-10-663-450-60	Sequence 60, Appl1
20	167.5	9.5	203	US-10-437-963-151695	Sequence 151695, A
21	167.5	9.5	203	US-10-732-923-13694	Sequence 13694, A
22	167.5	9.5	203	US-10-732-923-13695	Sequence 13695, A
23	164.5	9.3	176	US-10-437-963-164243	Sequence 164243, A
24	164.5	9.3	634	US-10-741-849-7083	Sequence 7083, Appl1
25	164	9.3	370	US-10-864-348-2	Sequence 2, Appl1
26	163.5	9.3	168	US-09-934-455-102	Sequence 102, Appl1
27	163.5	9.3	168	US-10-225-066A-206	Sequence 206, Appl1

ALIGNMENTS

RESULT 1

US-09-816-277-4
; Sequence 4, Application US/09816277
; Publication NO. US20010034045A1
; GENERAL INFORMATION:
; APPLICANT: Penttila, Merja E.
; APPLICANT: Ward, Michael
; APPLICANT: Wang, Huaming
; APPLICANT: Valkonen, Mari J.
; APPLICANT: Saloheimo, Markku
; TITLE OF INVENTION: Increased Production of Secreted
; TITLE OF INVENTION: Proteins by Recombinant Eukaryotic Cells
; FILE REFERENCE: GC590-2
; CURRENT APPLICATION NUMBER: US/09/816,277
; CURRENT FILING DATE: 2001-03-23
; PRIOR APPLICATION NUMBER: US 09/534,692
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 349
; TYPE: PRT
; ORGANISM: Aspergillus nidulans
US-09-816-277-4

Query Match	100.0%;	Score 1766;	DB 3;	Length 349;
Best Local Similarity	100.0%;	Pred. No. 3.2e-118;	Mismatches 0;	Indels 0;
Matches 349;	Conservative 0;			Gaps 0;
Qy	1	MKSADRFSPVKMEDAFANSPPTPSLEVPVLTSPADTSLRTKNVVAQTKEPKKPAKRRK 60		
Db	1	MKSADRFSPVKMEDAFANSPPTPSLEVPVLTSPADTSLRTKNVVAQTKEPKKPAKRRK 60		
Qy	61	SWGQLPVPKTNLPPRKAKTDEKEQRRIERVLNRAAAQTSRERKRLEMEKLESEKID 120		
Db	61	SWGQLPVPKTNLPPRKAKTDEKEQRRIERVLNRAAAQTSRERKRLEMEKLESEKID 120		
Qy	121	MEQOQNFLLORLAQMEANNRLSQVAQSAEVRGSRHSPTSSSPASVSPTLTPTLFKQ 180		
Db	121	MEQOQNFLLORLAQMEANNRLSQVAQSAEVRGSRHSPTSSSPASVSPTLTPTLFKQ 180		
Qy	181	EGDEYPLDRIPPTPSVTDYSPTLKPSLSIAESPDLTQHPAVSVGGLEGDESALTFLDLGA 240		
Db	181	EGDEYPLDRIPPTPSVTDYSPTLKPSLSIAESPDLTQHPAVSVGGLEGDESALTFLDLGA 240		
Qy	241	SIKHEPTHDLTAPLSDDDFRRLFNQDSSLSLEDGSAFADVLDSGDLSPFPDMSVD 300		
Db	241	SIKHEPTHDLTAPLSDDDFRRLFNQDSSLSLEDGSAFADVLDSGDLSPFPDMSVD 300		
Qy	301	FDEPVTLEDLQTNGLSDSASCKAASLQPSHGASTSRCDGGGIAAGSA 349		

Db 301 FDTPEVTLEDLEQTNGLSDSASCKAASLQPSHGASTSRCDGGGIAAGSA 349
|||||

RESULT 2
US-09-816-277-6
; Sequence 6, Application US/09816277
; Publication No. US20010034045A1
; GENERAL INFORMATION:
; APPLICANT: Penttila, Merja E.
; APPLICANT: Ward, Michael
; APPLICANT: Wang, Huaming
; APPLICANT: Valkonen, Mari J.
; APPLICANT: Saloheimo, Markku
; TITLE OF INVENTION: Increased Production of Secreted
; FILE REFERENCE: GC590-2
; CURRENT APPLICATION NUMBER: US/09/816,277
; CURRENT FILING DATE: 2001-03-23
; PRIOR APPLICATION NUMBER: US 09/534,692
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 349
; TYPE: PRT
; ORGANISM: Aspergillus nidulans
US-09-816-277-6

Query Match 100.0%; Score 1766; DB 3; Length 349;
Best Local Similarity 100.0%; Pred. No. 3.2e-118;
Matches 349; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKSADRFSPVKMEDAFANSPPTPSLEVPVLTVPADTSLRTKNVVAQTKEPKKPAKKR 60
Db 1 MKSADRFSPVKMEDAFANSPPTPSLEVPVLTVPADTSLRTKNVVAQTKEPKKPAKKR 60
QY 61 SWGQELPVPKTNLPKRKRAKTEDEKEQRIERVLNRNRAAQTSLRTKNVVAQTKEPKKLESEKID 120
Db 61 SWGQELPVPKTNLPKRKRAKTEDEKEQRIERVLNRNRAAQTSLRTKNVVAQTKEPKKLESEKID 120
QY 121 MEQONQFLLQRLAQMEAEENRLSQVAQLSAEVRGSRHSTPTSSSPASVSPTLTPTLTKQ 180
Db 121 MEQONQFLLQRLAQMEAEENRLSQVAQLSAEVRGSRHSTPTSSSPASVSPTLTPTLTKQ 180
QY 181 EGDEVPLDRIPPTPSVTDSPTLKPSSLAESPDLTQHPAVSVGGLEGDESALTJFDLGA 240
Db 181 EGDEVPLDRIPPTPSVTDSPTLKPSSLAESPDLTQHPAVSVGGLEGDESALTJFDLGA 240
QY 241 SIKHEPTHDLTAPLSDDDFRRLFNGDSSLESDDLLEDGFAFDVLDGSLSAFFPDSMVD 300
Db 241 SIKHEPTHDLTAPLSDDDFRRLFNGDSSLESDDLLEDGFAFDVLDGSLSAFFPDSMVD 300
QY 301 FDTPEVTLEDLEQTNGLSDSASCKAASLQPSHGASTSRCDGGGIAAGSA 349
Db 301 FDTPEVTLEDLEQTNGLSDSASCKAASLQPSHGASTSRCDGGGIAAGSA 349

RESULT 3
US-10-663-450-4
; Sequence 4, Application US/10663450
; Publication No. US20040186070A1
; GENERAL INFORMATION:
; APPLICANT: Penttila, Merja E.
; APPLICANT: Ward, Michael
; APPLICANT: Wang, Huaming
; APPLICANT: Valkonen, Mari J.
; APPLICANT: Saloheimo, Markku
; TITLE OF INVENTION: Increased Production of Secreted
; FILE REFERENCE: GC590-2
; CURRENT APPLICATION NUMBER: US/10/663,450
; CURRENT FILING DATE: 2003-09-15

; PRIOR APPLICATION NUMBER: US 09/534,692
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 349
; TYPE: PRT
; ORGANISM: Aspergillus nidulans
US-10-663-450-4

Query Match 100.0%; Score 1766; DB 4; Length 349;
Best Local Similarity 100.0%; Pred. No. 3.2e-118;
Matches 349; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKSADRFSPVKMEDAFANSPPTPSLEVPVLTVPADTSLRTKNVVAQTKEPKKPAKKR 60
Db 1 MKSADRFSPVKMEDAFANSPPTPSLEVPVLTVPADTSLRTKNVVAQTKEPKKPAKKR 60
QY 61 SWGQELPVPKTNLPKRKRAKTEDEKEQRIERVLNRNRAAQTSLRTKNVVAQTKEPKKLESEKID 120
Db 61 SWGQELPVPKTNLPKRKRAKTEDEKEQRIERVLNRNRAAQTSLRTKNVVAQTKEPKKLESEKID 120
QY 121 MEQONQFLLQRLAQMEAEENRLSQVAQLSAEVRGSRHSTPTSSSPASVSPTLTPTLTKQ 180
Db 121 MEQONQFLLQRLAQMEAEENRLSQVAQLSAEVRGSRHSTPTSSSPASVSPTLTPTLTKQ 180
QY 181 EGDEVPLDRIPPTPSVTDSPTLKPSSLAESPDLTQHPAVSVGGLEGDESALTJFDLGA 240
Db 181 EGDEVPLDRIPPTPSVTDSPTLKPSSLAESPDLTQHPAVSVGGLEGDESALTJFDLGA 240
QY 241 SIKHEPTHDLTAPLSDDDFRRLFNGDSSLESDDLLEDGFAFDVLDGSLSAFFPDSMVD 300
Db 241 SIKHEPTHDLTAPLSDDDFRRLFNGDSSLESDDLLEDGFAFDVLDGSLSAFFPDSMVD 300
QY 301 FDTPEVTLEDLEQTNGLSDSASCKAASLQPSHGASTSRCDGGGIAAGSA 349
Db 301 FDTPEVTLEDLEQTNGLSDSASCKAASLQPSHGASTSRCDGGGIAAGSA 349

RESULT 4
US-10-663-450-6
; Sequence 6, Application US/10663450
; Publication No. US20040186070A1
; GENERAL INFORMATION:
; APPLICANT: Penttila, Merja E.
; APPLICANT: Ward, Michael
; APPLICANT: Wang, Huaming
; APPLICANT: Valkonen, Mari J.
; APPLICANT: Saloheimo, Markku
; TITLE OF INVENTION: Increased Production of Secreted
; FILE REFERENCE: GC590-2
; CURRENT APPLICATION NUMBER: US/10/663,450
; CURRENT FILING DATE: 2003-09-15
; PRIOR APPLICATION NUMBER: US 09/534,692
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 349
; TYPE: PRT
; ORGANISM: Aspergillus nidulans
US-10-663-450-6

Query Match 100.0%; Score 1766; DB 4; Length 349;
Best Local Similarity 100.0%; Pred. No. 3.2e-118;
Matches 349; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKSADRFSPVKMEDAFANSPPTPSLEVPVLTVPADTSLRTKNVVAQTKEPKKPAKKR 60
Db 1 MKSADRFSPVKMEDAFANSPPTPSLEVPVLTVPADTSLRTKNVVAQTKEPKKPAKKR 60
QY 61 SWGQELPVPKTNLPKRKRAKTEDEKEQRIERVLNRNRAAQTSLRTKNVVAQTKEPKKLESEKID 120

```
Db 61 SWGQLPVKTNLPKRAKTEDEKQRRIRVLRNRAAAQTSRKRKLEMEKLESEKID 120
Qy 121 MEQONQFLQRLAQMEANNRLSQVAQLSAEVRGSRHSTPTSSSPASVSPTLTPTLFKQ 180
Db 121 MEQONQFLQRLAQMEANNRLSQVAQLSAEVRGSRHSTPTSSSPASVSPTLTPTLFKQ 180
Qy 181 EGVPLDRIPPTPTSPVTDYSPTLKPSSLAESPDLTQHPAVSVGGLEGDESALTLPDLGA 240
Db 181 EGVPLDRIPPTPTSPVTDYSPTLKPSSLAESPDLTQHPAVSVGGLEGDESALTLPDLGA 240
Qy 241 SIKHPTHDTAPLSDDDFRLFNGLSDSSLESLSLLEDGFAFDVLDGDLSAFPFDSMVD 300
Db 241 SIKHPTHDTAPLSDDDFRLFNGLSDSSLESLSLLEDGFAFDVLDGDLSAFPFDSMVD 300
Qy 301 FDTPEVTLEDQTNGLSDSASCKAASLPQSHGASTSRCDGGQIAAGSA 349
Db 301 FDTPEVTLEDQTNGLSDSASCKAASLPQSHGASTSRCDGGQIAAGSA 349
```

RESULT 5

US-09-816-277-19

; Sequence 19, Application US/09816277

; Publication No. US20010034045A1

; GENERAL INFORMATION:

; APPLICANT: Penttila, Merja E.

; APPLICANT: Ward, Michael

; APPLICANT: Wang, Huaming

; APPLICANT: Valkonen, Mari J.

; APPLICANT: Saloheimo, Markku

; TITLE OF INVENTION: Increased Production of Secreted

; TITLE OF INVENTION: Proteins by Recombinant Eukaryotic Cells

; FILE REFERENCE: GC590-2

; CURRENT APPLICATION NUMBER: US/09/816,277

; PRIOR FILING DATE: 2001-03-23

; PRIOR FILING DATE: 2000-03-24

; NUMBER OF SEQ ID NOS: 63

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 19

; LENGTH: 342

; TYPE: PRT

; ORGANISM: Aspergillus niger

US-09-816-277-19

```
Query Match 73.2%; Score 1293.5; DB 3; Length 342;
Best Local Similarity 76.2%; Pred. No. 2.1e-84;
Matches 262; Conservative 29; Mismatches 46; Indels 7; Gaps 3;

Qy 5 DRFSPVKMEDAFANSPPTTSPVLTSPADTSLRTKNVA-OTKPEKKPAKKKSWG 63
Db 4 EAFSPV---DSLGSPTP---ELPLLTSPADTSLDDSSVQAGETKAEKKPKKKSWG 57
Qy 64 QELPVKTNLPKRAKTEDEKQRRIRVLRNRAAAQTSRKRKLEMEKLESEKIDMEQ 123
Db 58 QELPVKTNLPKRAKTEDEKQRRIRVLRNRAAAQTSRKRKLEMEKLENEKIQMEQ 117
Qy 124 QNQLLQRLAQMEANNRLSQVAQLSAEVRGSRHSTPTSSSPASVSPTLTPTLFKQEGD 183
Db 118 QNQLLQRLSQMEANNRLNQVAQLSAEVRGSRGNTPKGSPVSAASPTLTPTLFKQERD 177
Qy 184 EYPLDRIPPTPTSPVTDYSPTLKPSSLAESPDLTQHPAVSVGGLEGDESALTLPDLGASIK 243
Db 178 EYPLERIPPTPTITDYSPTLSPSTLAESSDVTQHPAVSVAGLEGGSALSFLDVGSNPE 237
Qy 244 HEPTHDTAPLSDDDFRLFNGLSDSSLESLSLLEDGFAFDVLDGDLSAFPFDSMVD 303
Db 238 PHAADDLAAPLSDDDFRLFNVDSPVGSDDSVLEDDGFAFDVLDGDLSAFPFDSMVD 297
Qy 304 EYVTLLEDQTNGLSDSASCKAASLPQSHGASTSRCDGGQIAAG 347
Db 298 ESVGFEIGIEPHGLPDETSTQTSVQPSLGASTRCDGGQIAAG 341
```

[illegible]

QY 329 -----QPSHGASTSRCDGGGIAAG 347
DB 411 RELDLEIHPENQIPSRHSIQPQSGASHGCDGGIAGV 450

RESULT 10

US-09-816-277-5
; Sequence 5, Application US/09816277
; Publication No. US20010034045A1
; GENERAL INFORMATION:
; APPLICANT: Penttila, Merja E.
; APPLICANT: Ward, Michael
; APPLICANT: Valkonen, Mari J.
; APPLICANT: Saloheimo, Markku
; TITLE OF INVENTION: Increased Production of Secreted
; TITLE OF INVENTION: Proteins by Recombinant Eukaryotic Cells
; FILE REFERENCE: GC590-2
; CURRENT APPLICATION NUMBER: US/09/816,277
; CURRENT FILING DATE: 2001-03-23
; PRIOR APPLICATION NUMBER: US 09/534,692
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 451
; TYPE: PRT
; ORGANISM: Trichoderma reesei
US-09-816-277-5

Query Match 24.6%; Score 434; DB 3; Length 451;
Best Local Similarity 32.4%; Pred. No. 1e-22;
Matches 149; Conservative 50; Mismatches 109; Indels 152; Gaps 21;

QY 5 DRFSPVKMEDAFANSPPTPSLEVPVLTVPADTSLRTKNVQAQ-----47
DB 26 DNFTSL-----FADS--TPS-----TLNPRD--MMTPDSVADISRLSVIPESQDAEDD 70
QY 48 -----TKP--EKKPAKRRKSWGQELPVKTNLPKPKRAKTEDEKEQRRVERVLNR 98
DB 71 ESHSTSATAPSTSEKKPVKRRKSWGQVLPKPTNLPPKRAKTEDEKEQRRVERVLNR 130
QY 99 AAQTSRRKRLEMEKLESEK-----IDMEQONQFLQRLAQMBAENRNL 143
DB 131 AAQSSRRKRLEVEALEKRNKELETLINVOKTNLILVEELNRRFRSSGVVTRSSPLDS 190
QY 144 -QOVAQLSAEVRGSRH---SPTSS-----SPASVSPILTPTLFKQ---180
DB 191 LQDSITLSQQLFGSRDQGTMSNPEQSLMDQIMRSAANTVNPASLSPSPISDKKEFQTK 250
QY 181 EGDEVPLDRIPPTPSVTDSYPTLKPSLAE-----SPDLTQHPAVSVGGLGDESA 232
DB 251 EDEDEQADE--DEEMEQTWHTKEAAAKKNSKQSRVSTDTORPAVSIIG---DAA 303
QY 233 LTLP--DLGAS--IKHEPTHDLTAPLSDDDFRRLFNQDSSLESDDLLEGFADFVLDSDG 289
DB 304 VPFVSDDAGANCLGLDPVHQDDGPPS---IGHSFGLSALDADRYLLE-----SQ 350
QY 290 LSAPFPDSMVDFD-----TEPVTLLE-DLEQTNGLSDSASCKAASL-----328
DB 351 LLASFNASTVDDYLAGDSAACTNPLPSDYDFDINFLTDANHAAYDIVAASNYAAAD 410
QY 329 -----QPSHGASTSRCDGGGIAAG 347
DB 411 RELDLEIHPENQIPSRHSIQPQSGASHGCDGGIAGV 450

RESULT 11

US-10-663-450-2
; Sequence 2, Application US/10663450
; Publication No. US20040186070A1
; GENERAL INFORMATION:

; APPLICANT: Penttila, Merja E.
; APPLICANT: Ward, Michael
; APPLICANT: Valkonen, Mari J.
; APPLICANT: Saloheimo, Markku
; TITLE OF INVENTION: Increased Production of Secreted
; TITLE OF INVENTION: Proteins by Recombinant Eukaryotic Cells
; FILE REFERENCE: GC590-2
; CURRENT APPLICATION NUMBER: US/10/663,450
; CURRENT FILING DATE: 2003-09-15
; PRIOR APPLICATION NUMBER: US 09/534,692
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 451
; TYPE: PRT
; ORGANISM: Trichoderma reesei
US-10-663-450-2

Query Match 24.6%; Score 434; DB 4; Length 451;
Best Local Similarity 32.4%; Pred. No. 1e-22;
Matches 149; Conservative 50; Mismatches 109; Indels 152; Gaps 21;

QY 5 DRFSPVKMEDAFANSPPTPSLEVPVLTVPADTSLRTKNVQAQ-----47
DB 26 DNFTSL-----FADS--TPS-----TLNPRD--MMTPDSVADISRLSVIPESQDAEDD 70
QY 48 -----TKP--EKKPAKRRKSWGQELPVKTNLPKPKRAKTEDEKEQRRVERVLNR 98
DB 71 ESHSTSATAPSTSEKKPVKRRKSWGQVLPKPTNLPPKRAKTEDEKEQRRVERVLNR 130
QY 99 AAQTSRRKRLEMEKLESEK-----IDMEQONQFLQRLAQMBAENRNL 143
DB 131 AAQSSRRKRLEVEALEKRNKELETLINVOKTNLILVEELNRRFRSSGVVTRSSPLDS 190
QY 144 -QOVAQLSAEVRGSRH---SPTSS-----SPASVSPILTPTLFKQ---180
DB 191 LQDSITLSQQLFGSRDQGTMSNPEQSLMDQIMRSAANTVNPASLSPSPISDKKEFQTK 250
QY 181 EGDEVPLDRIPPTPSVTDSYPTLKPSLAE-----SPDLTQHPAVSVGGLGDESA 232
DB 251 EDEDEQADE--DEEMEQTWHTKEAAAKKNSKQSRVSTDTORPAVSIIG---DAA 303
QY 233 LTLP--DLGAS--IKHEPTHDLTAPLSDDDFRRLFNQDSSLESDDLLEGFADFVLDSDG 289
DB 304 VPFVSDDAGANCLGLDPVHQDDGPPS---IGHSFGLSALDADRYLLE-----SQ 350
QY 290 LSAPFPDSMVDFD-----TEPVTLLE-DLEQTNGLSDSASCKAASL-----328
DB 351 LLASFNASTVDDYLAGDSAACTNPLPSDYDFDINFLTDANHAAYDIVAASNYAAAD 410
QY 329 -----QPSHGASTSRCDGGGIAAG 347
DB 411 RELDLEIHPENQIPSRHSIQPQSGASHGCDGGIAGV 450

RESULT 12

US-10-663-450-5
; Sequence 5, Application US/10663450
; Publication No. US20040186070A1
; GENERAL INFORMATION:
; APPLICANT: Penttila, Merja E.
; APPLICANT: Ward, Michael
; APPLICANT: Valkonen, Mari J.
; APPLICANT: Saloheimo, Markku
; TITLE OF INVENTION: Increased Production of Secreted
; TITLE OF INVENTION: Proteins by Recombinant Eukaryotic Cells
; FILE REFERENCE: GC590-2
; CURRENT APPLICATION NUMBER: US/10/663,450
; CURRENT FILING DATE: 2003-09-15
; PRIOR APPLICATION NUMBER: US 09/534,692

```
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 451
; TYPE: PRT
; ORGANISM: Trichoderma reesei
US-10-663-450-5

Query Match      24.6%; Score 434; DB 4; Length 451;
Best Local Similarity 32.4%; Pred. No. 1e-22;
Matches 149; Conservative 50; Mismatches 109; Indels 152; Gaps 21;

QY 5 DRFSVKMEDAFANSTPTTSLEVPVLTVPADTSLRTKNVVAQ----- 47
Db 26 DNFTSL-----FADS--TPS-----TLNPRD--MWTDSVADIDSRLSVIPESQDAEDD 70
QY 48 -----TKP--EEKPAKAKRSWQOELPVPKTNLPKPKRAKTEDEKEQRIERVLRNRA 98
Db 71 ESHSTSATAPSTSEKKPKVKRKSWQVLPPEKTNLPKPKRAKTEDEKEQRRVERVLRNR 130
QY 99 AAOQSRERKRLMEKLESEK-----IDMEQOQFLQRLAQMEANRLS----- 143
Db 131 AAQSSRRERKLEVEALEKRNKELETLLINQVQKTNLILVEELNFRFRSSGVVTRSSPLDS 190
QY 144 -QOVAQLSAEVRGSRH-----STPTSS-----SPASVSTLTPTLTFKQ----- 180
Db 191 LQDSITLSQQLFGSDGQTMSPNQSLMDQIMRSAANTVNPASLSPSLPISDKFQTK 250
QY 181 EGDEVPRLRIPPTPSVTDYSSTLKPSLAE-----SPDLTOHPAVSVGGLEDSEA 232
Db 251 EDEEQADE---DEMEQWTHETKEAAAKENKSKQSRVSTDTQRPVAVSIGG---DAA 303
QY 233 LTLF--DIGAS-IKHEPTHDLTAPLSDDDFRLENGDSLESDDLLEDGFAFDVLDSDG 289
Db 304 VPFVSDADAGCLGLDPVHQDDGPFSS---IGHSFGLSAAALDADRYLLE-----SQ 350
QY 290 LSAPFSDSMVDFD-----TEPVTLLE-DLEQTNGLSDSASCKAASL----- 328
Db 351 LLASPNASTVDDYLLAGDSAACTWPLPSDYDFDINDFLTDDANHAAYDIVAASNVAAD 410
QY 329 -----QPSHGASTSRCDGOGIAG 347
Db 411 RELDLEIHDPENQIPSRHSIQPQSGASHGCDGDIAGV 450

RESULT 13
US-09-533-029-104
; Sequence 104, Application US/09533029
; Publication No. US20030046723A1
; GENERAL INFORMATION:
; APPLICANT: Heard, Jacqueline
; APPLICANT: Broun, Pierre
; APPLICANT: Riechmann, Jose-Luis
; APPLICANT: Keddie, James
; APPLICANT: Pineda, Omaira
; APPLICANT: Adam, Luc
; APPLICANT: Samaha, Raymond
; APPLICANT: Zhang, James
; APPLICANT: Yu, Guo-Liang
; APPLICANT: Ratcliffe, Oliver
; APPLICANT: Pilgrim, Marsha
; APPLICANT: Jiang, Cai-Zhong
; APPLICANT: Reuber, Lynne
; TITLE OF INVENTION: DISEASE-INDUCED POLYNUCLEOTIDES
; FILE REFERENCE: MBI-010
; CURRENT APPLICATION NUMBER: US/09/533,029
; EARLIER FILING DATE: 2000-03-22
; CURRENT FILING DATE: 2000-03-22
; EARLIER FILING DATE: 1999-03-23
; NUMBER OF SEQ ID NOS: 121
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 104
```

```
; LENGTH: 409
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; OTHER INFORMATION: G1034
US-09-533-029-104

Query Match      22.6%; Score 398.5; DB 3; Length 409;
Best Local Similarity 45.3%; Pred. No. 3.2e-20;
Matches 107; Conservative 26; Mismatches 84; Indels 19; Gaps 6;

QY 10 VKMEDAFANSTPTTSLEVPVLTVSPA--DTSRLTKNVVAOTKPEKKPAKRSWQOELP 67
Db 22 VSTPSSFNPLFNNLNPVDGFSQFSDRDYNNFNGSLGSLNLPKPKIKRKSJWGQQLP 81
QY 68 VPKTNLPKPKRAKTEDEKEQRIERVLRNRAAAQTSRERKLEMEKLESEKIDMEQQNOF 127
Db 82 EPKTNLPKPKRAKTEDEKEQRRVERVLRNRAAQSSRRERKQEVLEVEKRAIERKNWD 141
QY 128 LLQRLAQMEANRLSQQVAQLSAE-----VRGRHSTPTTSSSPASVSP-TLTPTLFKOE 181
Db 142 LEMRLADMEAKYVLLQQLKELKRAAGVYKTNFSLYSDSSTPDISEDSQLSPLTFSKQLFNAQ 201
QY 182 GDEVPDLRIP---PPTPSVTDYSSTLKPSL-----AESPDLTQHPAVSVGGLE 227
Db 202 -DELCRPISPOSIGELTSTRTVDPSTLSPKSLSSPDSSNSNSSDMTQHPAVVLCDLQ 256

RESULT 14
US-10-295-403-158
; Sequence 158, Application US/10295403
; Publication No. US20030101481A1
; GENERAL INFORMATION:
; APPLICANT: Heard, Jacqueline
; APPLICANT: Riechmann, Jose Luis
; APPLICANT: Adam, Luc
; APPLICANT: Broun, Pierre
; APPLICANT: Pineda, Omaira
; APPLICANT: Reuber, Lynne
; APPLICANT: Jiang, Cai-Zhong
; APPLICANT: Keddie, James
; APPLICANT: Zhang, James
; APPLICANT: Benito, Maria-Ines
; APPLICANT: Yu, Guo-Liang
; APPLICANT: Fromm, Mike
; TITLE OF INVENTION: PLANT GENE SEQUENCES I
; FILE REFERENCE: MBI-0003
; CURRENT APPLICATION NUMBER: US/10/295,403
; CURRENT FILING DATE: 2002-11-15
; PRIOR APPLICATION NUMBER: US/09/394,519
; PRIOR FILING DATE: 1999-09-13
; PRIOR APPLICATION NUMBER: 60/101,349
; PRIOR FILING DATE: 1998-09-22
; PRIOR APPLICATION NUMBER: 60/103,312
; PRIOR FILING DATE: 1998-10-06
; PRIOR APPLICATION NUMBER: 60/108,734
; PRIOR FILING DATE: 1998-11-17
; PRIOR APPLICATION NUMBER: 60/113,409
; PRIOR FILING DATE: 1998-12-22
; NUMBER OF SEQ ID NOS: 170
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 158
; LENGTH: 409
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; OTHER INFORMATION: G1034
US-10-295-403-158

Query Match      22.6%; Score 398.5; DB 4; Length 409;
Best Local Similarity 45.3%; Pred. No. 3.2e-20;
Matches 107; Conservative 26; Mismatches 84; Indels 19; Gaps 6;
```

QY 10 VKMEDAFANSPPTPSLEVPVLTVSPA--DTSLRTKNVVAQTKPEEKPAKRRKSWGQELP 67
Db 22 VSTPSSFHNPPLFDNLLNPNVDGFSQSFDRDYNFNGSLGLNLPKPKPIKRRKSWGQQLP 81
QY 68 VPKTNLPKRKRAKTEDEKEORRIERVLRNRAAAQTSRERKRLMEKLESEKIDMEQONQF 127
Db 82 EPKTNLPKRKRAKTQDEKEQRRVERVLRNRAAQSRRERKQVEALEVEKRAIERKNMD 141
QY 128 LLQRLAQMEANNRLSQVAQLSAE-----VRGSRHSTPTSSPASVSP-TLTPTTLFKOE 181
Db 142 LEMRLADMEAKYLLQQLKRAASGYNKTNPLSYSDSTPDISDSQSLPITFSKQLFNAQ 201
QY 182 GDEVPDLRIP---FPTPSVTDYSPTLKPSSL-----AESDPLTOHPAVSVGGLE 227
Db 202 -DELCPISQSIGPLTSTRVDPSTLSPKSLSPDSSNSNSDMDTQHPAVVLCDLQ 256

RESULT 15

US-10-412-699B-684
; Sequence 684, Application US/10412699B
; Publication No. US20040045049A1
; GENERAL INFORMATION:
; APPLICANT: Mendel Biotechnology, Inc.
; APPLICANT: Zhang, James
; APPLICANT: Fromm, Michael E.
; APPLICANT: Heard, Jacqueline E.
; APPLICANT: Riechmann, Jose Luis
; APPLICANT: Adam, Luc J.
; APPLICANT: Brown, Pierre E.
; APPLICANT: Pineda, Omaira
; APPLICANT: Reuber, T. Lynne
; APPLICANT: Keddle, James S.
; APPLICANT: Yu, Guo-Liang
; APPLICANT: Jiang, Cai-Zhong
; APPLICANT: Samaha, Raymond R.
; APPLICANT: Pilgrim, Marsha L.
; APPLICANT: Creelman, Robert A.
; APPLICANT: DuBell, Arnold N.
; APPLICANT: Ratcliffe, Oliver
; APPLICANT: Kumimoto, Roderick
; APPLICANT: Sherman, Bradley K.
; TITLE OF INVENTION: Polynucleotides and Polypeptides in Plants
; FILE REFERENCE: MBI-0048CIP
; CURRENT APPLICATION NUMBER: US/10/412,699B
; CURRENT FILING DATE: 2003-04-10
; PRIOR APPLICATION NUMBER: 09/394,519
; PRIOR FILING DATE: 1999-09-13
; PRIOR APPLICATION NUMBER: 09/489,376
; PRIOR FILING DATE: 2000-04-21
; PRIOR APPLICATION NUMBER: 09/506,720
; PRIOR FILING DATE: 2000-02-17
; PRIOR APPLICATION NUMBER: 09/533,030
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: 09/533,392
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: 09/533,029
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: 09/532,591
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: 09/533,648
; PRIOR FILING DATE: 2000-03-22
; PRIOR APPLICATION NUMBER: 09/713,994
; PRIOR FILING DATE: 2000-11-16
; PRIOR APPLICATION NUMBER: 09/819,142
; PRIOR FILING DATE: 2001-03-27
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 2011
; SOFTWARE: Patentin version 3.2
; SEQ ID NO 684
; LENGTH: 409
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:

; OTHER INFORMATION: G1034
US-10-412-699B-684

Query Match 22.6%; Score 398.5; DB 4; Length 409;
Best Local Similarity 45.3%; Pred. No. 3.2e-20;
Matches 107; Conservative 26; Mismatches 84; Indels 19; Gaps 6;
QY 10 VKMEDAFANSPPTPSLEVPVLTVSPA--DTSLRTKNVVAQTKPEEKPAKRRKSWGQELP 67
Db 22 VSTPSSFHNPPLFDNLLNPNVDGFSQSFDRDYNFNGSLGLNLPKPKPIKRRKSWGQQLP 81
QY 68 VPKTNLPKRKRAKTEDEKEORRIERVLRNRAAAQTSRERKRLMEKLESEKIDMEQONQF 127
Db 82 EPKTNLPKRKRAKTQDEKEQRRVERVLRNRAAQSRRERKQVEALEVEKRAIERKNMD 141
QY 128 LLQRLAQMEANNRLSQVAQLSAE-----VRGSRHSTPTSSPASVSP-TLTPTTLFKOE 181
Db 142 LEMRLADMEAKYLLQQLKRAASGYNKTNPLSYSDSTPDISDSQSLPITFSKQLFNAQ 201
QY 182 GDEVPDLRIP---FPTPSVTDYSPTLKPSSL-----AESDPLTOHPAVSVGGLE 227
Db 202 -DELCPISQSIGPLTSTRVDPSTLSPKSLSPDSSNSNSDMDTQHPAVVLCDLQ 256

Search completed: November 23, 2005, 03:32:21
Job time : 114.983 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 23, 2005, 03:12:51 ; Search time 34.1763 Seconds
(without alignments)
844.265 Million cell updates/sec

Title: US-10-663-450-4
Perfect score: 1766
Sequence: 1 MKSADRFSPVKMEDAFANSP.....PSHGASTSRCDGGIAGSA 349

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:*

- 1: /cgn2_6/ptodata/1/iaa/5_COMB.pep.*
- 2: /cgn2_6/ptodata/1/iaa/6_COMB.pep.*
- 3: /cgn2_6/ptodata/1/iaa/H_COMB.pep.*
- 4: /cgn2_6/ptodata/1/iaa/PCTUS_COMB.pep.*
- 5: /cgn2_6/ptodata/1/iaa/RE_COMB.pep.*
- 6: /cgn2_6/ptodata/1/iaa/backfilese1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	398.5	22.6	409	2	US-09-533-029-104
2	194	11.0	248	2	US-09-248-796A-18840
3	160	9.1	264	2	US-09-949-016-10789
4	159.5	9.0	260	2	US-09-538-092-950
5	145.5	8.2	1402	2	US-09-248-796A-14503
6	141.5	8.0	521	1	US-08-721-684C-2
7	141.5	8.0	521	1	US-09-005-970-2
8	141.5	8.0	521	2	US-09-407-715-2
9	141	8.0	340	2	US-09-350-841A-1601
10	137.5	7.8	143	2	US-09-840-211A-1930
11	135.5	7.7	395	2	US-09-247-155-113
12	135.5	7.7	395	2	US-09-513-999C-14
13	135.5	7.7	395	2	US-09-471-276-14
14	135.5	7.7	395	2	US-09-903-190-113
15	134.5	7.6	380	2	US-09-165-522-16
16	133	7.5	102	2	US-09-840-211A-1046
17	133	7.5	338	1	US-08-318-686-2
18	133	7.5	338	2	US-08-460-242-2
19	133	7.5	1341	2	US-09-949-016-6890
20	133	7.5	1344	2	US-09-949-016-10925
21	132.5	7.5	380	2	US-08-486-099-115
22	132.5	7.5	380	2	US-08-484-223B-115
23	132.5	7.5	380	2	US-08-919-597-115
24	132.5	7.5	380	2	US-08-475-668A-115
25	132.5	7.5	380	2	US-08-485-551A-115
26	132.5	7.5	380	2	US-08-471-913A-115
27	132.5	7.5	380	2	US-08-485-264A-115

Sequence 115, App
Sequence 115, App
Sequence 115, App
Sequence 115, App
Sequence 115, App
Sequence 24, Appl
Sequence 24, Appl
Sequence 24, Appl
Sequence 125, App
Sequence 125, App
Sequence 4, Appli
Sequence 4, Appli
Sequence 4, Appli
Sequence 22, Appl
Sequence 22, Appl
Sequence 14492, A
Sequence 9676, Ap

ALIGNMENTS

RESULT 1

US-09-533-029-104
; Sequence 104, Application US/09533029
; Patent No. 6664446
; GENERAL INFORMATION:
; APPLICANT: Heard, Jacqueline
; APPLICANT: Broun, Pierre
; APPLICANT: Riechmann, Jose-Luis
; APPLICANT: Keddle, James
; APPLICANT: Pineda, Omaira
; APPLICANT: Adam, Luc
; APPLICANT: Samaha, Raymond
; APPLICANT: Zhang, James
; APPLICANT: Yu, Guo-liang
; APPLICANT: Ratcliffe, Oliver
; APPLICANT: Pilgrim, Marsha
; APPLICANT: Jiang, Cai-Zhong
; APPLICANT: Reuber, Lynne
; TITLE OF INVENTION: DISEASE-INDUCED POLYNUCLEOTIDES
; FILE REFERENCE: MBI-010
; CURRENT APPLICATION NUMBER: US/09/533,029
; CURRENT FILING DATE: 2000-03-22
; EARLIER APPLICATION NUMBER: 60/125,814
; EARLIER FILING DATE: 1999-03-23
; NUMBER OF SEQ ID NOS: 121
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 104
; LENGTH: 409
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; OTHER INFORMATION: G1034
US-09-533-029-104

Query Match 22.6%; Score 398.5; DB 2; Length 409;
Best Local Similarity 45.3%; Pred. No. 8.8e-29;
Matches 107; Conservative 26; Mismatches 84; Indels 19; Gaps 6;
QY 10 VKMEDAFANSPPTSPLEVPVLTVSPA--DTSURTKNVVAQTPEEKPAKRSWGQELP 67
Db 22 VSTPSFTHNPPLFDNNLPVDFGSPQSFDRDYNFNGSLSLGLNLPEKKPKKRSWGQQLP 81
QY 68 VPKTNLPKRKAKTDEKEORIERVLNRRAAQTSSRERKLEMEKLESEKIDMEQONQF 127
Db 82 EPTKTNLPKRKAKTQDEKEQRVERVLNRRAAQTSSRERKLEMEKLESEKIDMEQONQF 141
QY 128 LIQRLAOMEAENNRSLQQVAQLASAE-----VGRSRHSTPTSSSPASVSP-TLTPTLFKOE 181
Db 142 LEWRLADMEAKYLLQQLKRAAGYKNTNFTLSYSDSTPDISEDSQLSPITFSKQLFNAQ 201


```
/
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Williams Jr., Joseph A.
/ REGISTRATION NUMBER: 38,659
/ REFERENCE/DOCKET NUMBER: 27866/33487
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 312-474-6300
/ TELEFAX: 312-474-0448
/ INFORMATION FOR SEQ ID NO: 2:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 521 amino acids
/ TYPE: amino acid
/ TOPOLOGY: linear
/ MOLECULE TYPE: protein
/ SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-005-970-2

Query Match      8.0%; Score 141.5; DB 1; Length 521;
Best Local Similarity 24.0%; Pred. No. 0.00015;
Matches 82; Conservative 48; Mismatches 131; Indels 81; Gaps 16;

QY 8 SPVKMEDAFANSPPTPSLEVPVLTVPSPADTSLRTKNVVAQTKPEKKPAKREKSWGOELP 67
DB 234 SPVR---AMPRGPSALSTS-PLLT---APHKLGSGPLVLTTEE-----KKTIVAEGYP 280
QY 68 VPKNLPPRKRAKTEDEKEQRIIRIERNRAAAQTSRRKRLEMEKLESEKIDMQNQF 127
DB 281 IP-TKLPLTK---SEKALKKIRKIKNKISAQESRRKKKEYMDSLEKKVESCSTENLE 335
QY 128 LLQRLAQMENAENRLSQVAQLSAEVRGSRHSTPTSSPASVSPILTP-----176
DB 336 LRKKVEVLNRTLLQQLKQLTLVMGK-----VSRCKLAGTGTGCLMVMVVLCAV 389
QY 177 ----LFKQEGDEVPLDRIPFPPTPSVTDSPTLKPSSLAESPDLTQHPAVSVGLEGDESA 232
DB 390 AFGSFFQGYG-----PYPSATKMA---LPS-----QHP-----LSEPYTA 421
QY 233 LTLFDLGASI--KHEPTHDLTAPLDDDDFRRLFNQDSSLESDDLLEDGAFDVLDSGDL 290
DB 422 SVVRSRNLIIYEEHAPLESSSPASTGELGGWDGSSLLRASSGL-----EALPEVDL 474
QY 291 SAFFPDSMVDPTEPTVLTLEDLEQ---TNGLSDSASCKAASLQ 329
DB 475 PHFLISN--ETSLEKSVLLEQLQHLVSSKLEGNETLKVVELE 514

RESULT 8
US-09-407-715-2
/ Sequence 2, Application US/09407715
/ Patent No. 6248532
/ GENERAL INFORMATION:
/ APPLICANT: Keegan, Kathleen S.
/ TITLE OF INVENTION: No. 6248532e1 CREBa Isoform
/ NUMBER OF SEQUENCES: 10
/ CORRESPONDENCE ADDRESS:
/ ADDRESSER: Marshall, O'Toole, Gerstein, Murray & Borun
/ STREET: 233 South Wacker Drive, 6300 Sears Tower
/ CITY: Chicago
/ STATE: Illinois
/ COUNTRY: USA
/ ZIP: 60606
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: PatentIn Release #1.0, Version #1.30
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/09/407,715
/ FILING DATE: 28-Sep-1999
/ PRIOR APPLICATION DATA:
/ APPLICATION NUMBER: US/09/005,970
/ FILING DATE: 12-JAN-1998
/ APPLICATION NUMBER: US/08/721,684
/ FILING DATE: 27-SEP-1996
/ ATTORNEY/AGENT INFORMATION:
```

```
/
/ NAME: Williams Jr., Joseph A.
/ REGISTRATION NUMBER: 38,659
/ REFERENCE/DOCKET NUMBER: 27866/33487
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: 312-474-6300
/ TELEFAX: 312-474-0448
/ INFORMATION FOR SEQ ID NO: 2:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 521 amino acids
/ TYPE: amino acid
/ TOPOLOGY: linear
/ MOLECULE TYPE: protein
/ SEQUENCE DESCRIPTION: SEQ ID NO: 2:
US-09-407-715-2

Query Match      8.0%; Score 141.5; DB 2; Length 521;
Best Local Similarity 24.0%; Pred. No. 0.00015;
Matches 82; Conservative 48; Mismatches 131; Indels 81; Gaps 16;

QY 8 SPVKMEDAFANSPPTPSLEVPVLTVPSPADTSLRTKNVVAQTKPEKKPAKREKSWGOELP 67
DB 234 SPVR---AMPRGPSALSTS-PLLT---APHKLGSGPLVLTTEE-----KKTIVAEGYP 280
QY 68 VPKNLPPRKRAKTEDEKEQRIIRIERNRAAAQTSRRKRLEMEKLESEKIDMQNQF 127
DB 281 IP-TKLPLTK---SEKALKKIRKIKNKISAQESRRKKKEYMDSLEKKVESCSTENLE 335
QY 128 LLQRLAQMENAENRLSQVAQLSAEVRGSRHSTPTSSPASVSPILTP-----176
DB 336 LRKKVEVLNRTLLQQLKQLTLVMGK-----VSRCKLAGTGTGCLMVMVVLCAV 389
QY 177 ----LFKQEGDEVPLDRIPFPPTPSVTDSPTLKPSSLAESPDLTQHPAVSVGLEGDESA 232
DB 390 AFGSFFQGYG-----PYPSATKMA---LPS-----QHP-----LSEPYTA 421
QY 233 LTLFDLGASI--KHEPTHDLTAPLDDDDFRRLFNQDSSLESDDLLEDGAFDVLDSGDL 290
DB 422 SVVRSRNLIIYEEHAPLESSSPASTGELGGWDGSSLLRASSGL-----EALPEVDL 474
QY 291 SAFFPDSMVDPTEPTVLTLEDLEQ---TNGLSDSASCKAASLQ 329
DB 475 PHFLISN--ETSLEKSVLLEQLQHLVSSKLEGNETLKVVELE 514

RESULT 9
US-09-350-841A-1601
/ Sequence 1601, Application US/09350841A
/ Patent No. 6750008
/ GENERAL INFORMATION:
/ APPLICANT: Jeffs, Peter;
/ TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR INHIBITION OF MEMBRANE
/ TITLE OF INVENTION: FUSION-ASSOCIATED EVENTS, INCLUDING HIV TRANSMISSION
/ FILE REFERENCE: 7872-066-999
/ CURRENT APPLICATION NUMBER: US/09/350,841A
/ CURRENT FILING DATE: 1999-07-09
/ NUMBER OF SEQ ID NOS: 1946
/ SOFTWARE: PatentIn Ver. 2.1
/ SEQ ID NO 1601
/ LENGTH: 340
/ TYPE: PRT
/ ORGANISM: Human immunodeficiency virus type 1
US-09-350-841A-1601

Query Match      8.0%; Score 141; DB 2; Length 340;
Best Local Similarity 23.1%; Pred. No. 8.5e-05;
Matches 80; Conservative 40; Mismatches 12; Indels 100; Gaps 13;

QY 4 ADRF----SPVKMEDAFANSPPTPSLEVPVLTVPSPADTSLRTKNVVAQTKPEKKPAKCR 59
DB 35 ADSFSSMGSPVNAQDFCTDLAVSSANFIP--TVTAISTSPDLQWLQVAPALVSSVAPSQTR 92
QY 60 KSWGQELPVKTNLPPR-----KKAETE---DEKEQRIIRIERNRAA 100
```



```
Db 93 APHPFCVPAPSAGAYBAGVVKWTGCGRAQSIGRRKVKVEQLSEBEKKRRIRRE-RNKMAA 151
Qy 101 QTSREKRRLMEKLESEKIDMEQOQFLQRLQAEAEENRLS-----143
Db 152 AKCRNRRLTDTLOAETQLEDEKALQTEIANLLKEKEKLEFILAHRPACKIPDDLQ 211
Qy 144 --QOVAQLAEVRGSRHSTPTSSPASVSTLPTLPKQEGDEVLDRIPTPTPSVTDYS 201
Db 212 PFEEMSVASLDLFG---GLPEVATPS-----BEAFTPLJNDPEPKPSV---E 254
Qy 202 PTLKPSL---AESPDLOHPAVSGGLEGDESALTFLDGLASIKHEPTHDLTAPLS--- 255
Db 255 PVKSISSMELKTEPPDPSPVPEVKSISSME-----LKTEPDDFLFPASSRP 300
Qy 256 -----DDFRRLFGDSLSLSDSSLLSDGDFAFDVLDSGLSA 292
Db 301 SGSETARSVPDMDLGSFGVAGSSNBPSS-----DSLSPPTLIA 339

RESULT 10
US-09-640-211A-1930
; Sequence 1930, Application US/09640211A
; Patent No. 6833446
; GENERAL INFORMATION:
; APPLICANT: Wood, Marion
; APPLICANT: Shenk, Michael A.
; APPLICANT: McGrath, Annette
; APPLICANT: Glenn, Matthew
; TITLE OF INVENTION: Compositions and Methods for the
; TITLE OF INVENTION: Modification of Gene Transcription
; FILE REFERENCE: 11000.1021C1U
; CURRENT APPLICATION NUMBER: US/09/640,211A
; CURRENT FILING DATE: 2000-08-16
; NUMBER OF SEQ ID NOS: 2368
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1930
; LENGTH: 143
; TYPE: PRT
; ORGANISM: Pinus radiata
US-09-640-211A-1930

Query Match 7.8%; Score 137.5; DB 2; Length 143;
Best Local Similarity 24.8%; Pred. No. 4.8e-05;
Matches 38; Conservative 25; Mismatches 45; Indels 45; Gaps 2;

Qy 18 NSPTTPSLEVPVLTSPADTSLRTKNVVAQTKPEKKPAKRKSGWQELPVKPTNLPPRK 77
Db 5 SSVSPSPDTPSQSAARVPTSTRDSSVME-----PPRK 38
Qy 78 RAKTEDEKQRIERVLRNRAAQTSSRERKLEMEKLESEKIDMEQOQFL-----128
Db 39 RARADLNARQREARAHNRNRIAAQNSRDKRKAQFTYMEQORVAQLEENQRLRAGMGLSQF 98
Qy 129 -----LORLAQMEENRLSQOVAQLSA 151
Db 99 TPADNDKFVSLERESVQARENRELKERIKSLES 131

RESULT 11
US-09-247-155-113
; Sequence 113, Application US/09247155A
; Patent No. 6312922
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, Jean-Baptiste
; APPLICANT: Duclert, Aymeric
; APPLICANT: Bouqueleret, Lydie
; TITLE OF INVENTION: Complementary DNAs
; FILE REFERENCE: GENSET-021A
; CURRENT APPLICATION NUMBER: US/09/247,155A
; CURRENT FILING DATE: 1999-02-09
; EARLIER APPLICATION NUMBER: 60/074,121
; EARLIER FILING DATE: 1998-02-09
; EARLIER APPLICATION NUMBER: 60/081,563
```

```
; EARLIER FILING DATE: 1998-04-13
; EARLIER APPLICATION NUMBER: 60/096,116
; EARLIER FILING DATE: 1998-08-10
; EARLIER APPLICATION NUMBER: 60/099,273
; EARLIER FILING DATE: 1998-10-04
; NUMBER OF SEQ ID NOS: 182
; SOFTWARE: Patent.pm
; SEQ ID NO 113
; LENGTH: 395
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SIGNAL
; LOCATION: -310...-1
US-09-247-155-113

Query Match 7.7%; Score 135.5; DB 2; Length 395;
Best Local Similarity 25.2%; Pred. No. 0.00035;
Matches 62; Conservative 38; Mismatches 105; Indels 41; Gaps 10;

Qy 27 VEVLTVPADTSLRTKNVVAQTKPEKKPAKRKSGWQELPVKPTNLPPRKRAKTEDEKE 86
Db 176 VPCTTLLPCQTLFLT-----DEEK-----RLGQE-----GVSLPSHLPLTKAERV 217
Qy 87 QRIERVLRNRAAQTSSRERKLEMEKLESEKIDMEQOQFLQRLQAEAEENRLSQOV 146
Db 218 LKVRKRKIRNKQSAQDSRRRKEYIDGLERSVAAACSAQNOELQKVQELERHNISLVAOL 277
Qy 147 AQLSAEVRGSRHSTPTSSPASVSTLPTLPKQEGDEVLDRIPT---PTPSVTDYSP- 202
Db 278 RQLQTLI-----AQTSNKAQTSICVILLFSLALIILP-SFSPQSPRPEAGSEDYQPH 330
Qy 203 TLKPSSLAESPDLOHPAVSV-----GGLEGDESALTFLD-LGASIKHEPTHDLTA 252
Db 331 GVTSRNLTHTKQVTENLTQVVESRLRBPFGAKDANGSTRITLLEKMG--KRPSPGRIRS 388
Qy 253 PLSDDD 258
Db 389 VLHADE 394

RESULT 12
US-09-513-999C-14
; Sequence 14, Application US/09513999C
; Patent No. 6783961
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Duclert, A.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
; FILE REFERENCE: 59.US2.REG
; CURRENT APPLICATION NUMBER: US/09/513,999C
; CURRENT FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/122,487
; PRIOR FILING DATE: 1999-02-26
; NUMBER OF SEQ ID NOS: 36681
; SOFTWARE: Patent.pm
; SEQ ID NO 14
; LENGTH: 395
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SIGNAL
; LOCATION: -310...-1
; OTHER INFORMATION: score 4.4
; OTHER INFORMATION: seq VLIILFSLALIIL/PS
US-09-513-999C-14

Query Match 7.7%; Score 135.5; DB 2; Length 395;
Best Local Similarity 25.2%; Pred. No. 0.00035;
Matches 62; Conservative 38; Mismatches 105; Indels 41; Gaps 10;
```

Qy 27 VPVLTVSPADTSLRTKNVVAQTKEPKPAKRSWGQELPVPKTNLPKRKAKTEDEKE 86
Db 176 VPCITLLPCQTLFLT-----DEEK-----RLLGQE-----GVSLPSHLPLTKAERV 217

Qy 87 QRRIRVLNRNAAQTSSSRERKLEMEKLESEKIDMEQNOQFLQRLAOMEAENNRLSQOV 146
Db 218 LKKVRKIRNKQSAQDSRRRKKKEYIDGLESRVAACSAQNOELQKKVQELERHNISLVAQL 277

Qy 147 AQLSAEVRGSRHSTPTSSSPASVPTLTPTTLFKQEGDEVPLDRIPF---PTPSVTDYSP- 202
Db 278 RQLOTLI-----AQTSNKAAQTSTCVLILFLSLALIILP-SFSPFQSRPEAGSEDYQPH 330

Qy 203 TLKPSSLAESPDLTQHPAVSV-----GGLEGDESALTFLD-LGASIKHEPTHDLTA 252
Db 331 GVTSRNLTBKDVNTENLETQVVESRLRPPGAKDANGSTRTLLEKMGG--KPRPSGRIRS 388

Qy 253 PLSDDD 258
Db 389 VLHADE 394

RESULT 13
US-09-471-276-14
; Sequence 14, Application US/09471276
; Patent No. 6822072
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Duclert A.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
; FILE REFERENCE: GENSET.025CP1
; CURRENT APPLICATION NUMBER: US/09/471,276
; CURRENT FILING DATE: 1999-12-21
; EARLIER APPLICATION NUMBER: 09/057,719
; EARLIER FILING DATE: 1998-04-09
; EARLIER APPLICATION NUMBER: 09/069,047
; EARLIER FILING DATE: 1998-04-28
; EARLIER APPLICATION NUMBER: PCT/IB99/00712
; EARLIER FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 1622
; SOFTWARE: Patent.pm
; SEQ ID NO 14
; LENGTH: 395
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SIGNAL
; LOCATION: -310...-1
; OTHER INFORMATION: seq VLILFLSLALIIL/PS
US-09-471-276-14

Query Match 7.7%; Score 135.5; DB 2; Length 395;
Best Local Similarity 25.2%; Pred. No. 0.00035;
Matches 62; Conservative 38; Mismatches 105; Indels 41; Gaps 10;

Qy 27 VPVLTVSPADTSLRTKNVVAQTKEPKPAKRSWGQELPVPKTNLPKRKAKTEDEKE 86
Db 176 VPCITLLPCQTLFLT-----DEEK-----RLLGQE-----GVSLPSHLPLTKAERV 217

Qy 87 QRRIRVLNRNAAQTSSSRERKLEMEKLESEKIDMEQNOQFLQRLAOMEAENNRLSQOV 146
Db 218 LKKVRKIRNKQSAQDSRRRKKKEYIDGLESRVAACSAQNOELQKKVQELERHNISLVAQL 277

Qy 147 AQLSAEVRGSRHSTPTSSSPASVPTLTPTTLFKQEGDEVPLDRIPF---PTPSVTDYSP- 202
Db 278 RQLOTLI-----AQTSNKAAQTSTCVLILFLSLALIILP-SFSPFQSRPEAGSEDYQPH 330

Qy 203 TLKPSSLAESPDLTQHPAVSV-----GGLEGDESALTFLD-LGASIKHEPTHDLTA 252
Db 331 GVTSRNLTBKDVNTENLETQVVESRLRPPGAKDANGSTRTLLEKMGG--KPRPSGRIRS 388

Qy 253 PLSDDD 258
Db 389 VLHADE 394

RESULT 14
US-09-903-190-113
; Sequence 113, Application US/09903190
; Patent No. 6936692
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, Jean-Baptiste
; APPLICANT: Duclert, Aymeric
; APPLICANT: Bougueleret, Lydie
; TITLE OF INVENTION: Complementary DNAs
; FILE REFERENCE: GENSET.021A
; CURRENT APPLICATION NUMBER: US/09/903,190
; CURRENT FILING DATE: 2001-07-11
; PRIOR APPLICATION NUMBER: US/09/247,155A
; PRIOR FILING DATE: 1999-02-09
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/074,121
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-02-09
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/081,563
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/096,116
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-08-10
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/099,273
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-10-04
; NUMBER OF SEQ ID NOS: 182
; SOFTWARE: Patent.pm
; SEQ ID NO 113
; LENGTH: 395
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SIGNAL
; LOCATION: -310...-1
US-09-903-190-113

Query Match 7.7%; Score 135.5; DB 2; Length 395;
Best Local Similarity 25.2%; Pred. No. 0.00035;
Matches 62; Conservative 38; Mismatches 105; Indels 41; Gaps 10;

Qy 27 VPVLTVSPADTSLRTKNVVAQTKEPKPAKRSWGQELPVPKTNLPKRKAKTEDEKE 86
Db 176 VPCITLLPCQTLFLT-----DEEK-----RLLGQE-----GVSLPSHLPLTKAERV 217

Qy 87 QRRIRVLNRNAAQTSSSRERKLEMEKLESEKIDMEQNOQFLQRLAOMEAENNRLSQOV 146
Db 218 LKKVRKIRNKQSAQDSRRRKKKEYIDGLESRVAACSAQNOELQKKVQELERHNISLVAQL 277

Qy 147 AQLSAEVRGSRHSTPTSSSPASVPTLTPTTLFKQEGDEVPLDRIPF---PTPSVTDYSP- 202
Db 278 RQLOTLI-----AQTSNKAAQTSTCVLILFLSLALIILP-SFSPFQSRPEAGSEDYQPH 330

Qy 203 TLKPSSLAESPDLTQHPAVSV-----GGLEGDESALTFLD-LGASIKHEPTHDLTA 252
Db 331 GVTSRNLTBKDVNTENLETQVVESRLRPPGAKDANGSTRTLLEKMGG--KPRPSGRIRS 388

Qy 253 PLSDDD 258
Db 389 VLHADE 394

RESULT 15
US-09-165-522-16
; Sequence 16, Application US/09165522
; Patent No. 6943000
; GENERAL INFORMATION:
; APPLICANT: Davis, Roger J.
; Flavel, Richard A.
; Rakic, Pasko
; Whitmarsh, Alan
; Kuan, Chia-Yi

Yang, Di
TITLE OF INVENTION: JNK3 MODULATORS AND METHODS OF USE
NUMBER OF SEQUENCES: 24
CORRESPONDENCE ADDRESS:
ADDRESSEE: Fish & Richardson P.C.
STREET: 225 Franklin Street
CITY: Boston
STATE: MA
COUNTRY: USA
ZIP: 02110-2804
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows 95
SOFTWARE: FASTSEQ for Windows Version 2.0b
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/165,522
FILING DATE: 02-Oct-1998
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/060,995
FILING DATE: 03-OCT-1997
ATTORNEY/AGENT INFORMATION:
NAME: Faese, J. Peter
REGISTRATION NUMBER: 32,983
REFERENCE/DOCKET NUMBER: 10363/005001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617/542-5070
TELEFAX: 617/542-8906
TELEX: 200154
INFORMATION FOR SEQ ID NO: 16:
SEQUENCE CHARACTERISTICS:
LENGTH: 380 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 16:
US-09-165-522-16

Query Match 7.6%; Score 134.5; DB 2; Length 380;
Best Local Similarity 22.7%; Pred. No. 0.00041;
Matches 91; Conservative 48; Mismatches 135; Indels 127; Gaps 18;

Qy	4	ADRF---	SPVKMEDAFANSTPTPSLEVPVLTSPADTSLRTKNVVAQTKPEEKPAKRR	59
Db	34	ADSPSSMGSPVNTQFCADLSVSNANFIP--TVTAISTSPDLQWLQVPTLVSSVAPSQTR	91	
Qy	60	KSWGQLPVPKTNLPFR-----KRAKTE----	DEKQORRIERVLRNRAA	99
Db	92	APHYGLPTQAGAYARAGVMTVSGGQAQSIGRGKVEQLSPSEEEKRIRRE-RNKWA	150	
Qy	100	AQTSRERKRLMEKLESKIDMEQQNQFLQRLAQMEAEANNRLS-----	143	
Db	151	AAKCNRRRLTDTLQAEVDQLEDEKSAIQTEIANLLKEKEKLEFI LAHRPAKIPDDL	210	
Qy	144	---QQVAQLSAEVRGSRHSTPTSSPASVSPILTTLFKQEGDEVPLDRIPPTPSVTDY	200	
Db	211	GFPEEMSVASLDLGT---GLPEASTPES-----EFAFTPLPLNDPEPKPSL---	253	
Qy	201	SPTLKPSLSAESPDLTQHPAVSVGLEGDEGALTLFDLGASIKHEPTHDLTAPLSDDDFR	260	
Db	254	-----EPVKSINVE-----LKAEPFDDFLFPASSRP--	280	
Qy	261	RLFNGDSSLES--DSSLLEDGFADF--VLDSGLSAPFPDSMVDFDTPVTLEDLEQTNG	316	
Db	281	---SGSETSRSPVDVLDLGSFYAADWEPLHNSLGMGP---MVT-ELEPLCTPVVVTCTPG	333	
Qy	317	L-----SDS-ASCKAA-----SLQPSHGASTS	337	
Db	334	CTTYTSSPVFTYPEADSPFCAAAHRKGGSSNEPSSDSLSS	374	

Search completed: November 23, 2005, 03:27:33
Job time : 35.1763 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 23, 2005, 03:05:10 ; Search time 149.973 Seconds
(without alignments)
1641.819 Million cell updates/sec

Title: US-10-663-450-4

Perfect score: 1765

Sequence: 1 MKSADRFSPVKMEDAFANSP.....PSHGASTSRCDGQGIAGSA 349

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Uniprot 05.80.*

1: uniprot_sprot.*

2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	1755.5	99.4	350	Q8TFU8_EMENI	Q8tfu8 emericella
2	1696	96.0	347	Q5AQN3_EMENI	Q5agn3 aspergillus
3	1293.5	73.2	342	Q6W8X2_ASPNG	Q6w8x2 aspergillus
4	856	48.5	433	Q4WEY8_ASPFU	Q4wey8 aspergillus
5	434	24.6	451	Q8TFP3_TRIRE	Q8tfp3 trichoderma
6	423	24.0	430	Q7SHF0_NEUCR	Q7shf0 neurospora
7	412.5	23.4	429	Q4HT5_GIBZE	Q4ht5 gibberella
8	346	19.6	556	Q5IKW8_MAGGR	Q5ikw8 magnaporthe
9	261.5	14.8	299	Q6CEV1_YARLI	Q6cevl yarrowia li
10	217	12.3	273	Q6CKQ1_KJULA	Q6ckq1 kluyveromyc
11	209.5	11.9	260	Q6BQC2_DEBHA	Q6bcq2 debaryomyce
12	208	11.8	357	Q5AA52_CANAL	Q5aa52 candida alb
13	206.5	11.7	230	1 HAC1_YEAST	P41546 saccharomyc
14	177.5	10.1	396	Q7ZYC2_XENLA	Q7zyc2 xenopus lae
15	169	9.6	228	Q75BQ5_ASHGO	Q75bq5 ashbya goss
16	169	9.6	321	Q8GRY7_LOTJA	Q8gry7 lotus japon
17	168.5	9.5	350	Q90ZR7_XENLA	Q90zr7 xenopus lae
18	167.5	9.5	336	Q5VR11_ORYSA	Q5vr11 oryza sativ
19	166.5	9.4	437	Q5G9T3_XENLA	Q5g9t3 xenopus lae
20	166	9.4	458	Q4P9B7_USTMA	Q4p9b7 ustilago ma
21	166	9.4	639	Q6CET1_YARLI	Q6cet1 yarrowia li
22	165.5	9.4	322	Q39896_SOYBN	Q39896 glycine max
23	164.5	9.3	176	Q69XK6_ORYSA	Q69xk6 oryza sativ
24	164.5	9.3	634	Q59W43_CANAL	Q59w43 candida alb
25	164.5	9.3	634	Q59W83_CANAL	Q59w83 candida alb
26	163.5	9.3	168	1 HIS_ARATH	O24646 arabidopsis
27	163.5	9.3	470	Q5FVM5_RAT	Q5fvm5 rattus norv
28	163	9.2	252	Q6NX18_XENTR	Q6nx18 xenopus tro
29	163	9.2	326	Q39895_SOYBN	Q39895 glycine max
30	162	9.2	322	Q04234_VICFA	Q04234 vicia faba
31	161.5	9.1	329	Q6FLY3_CANGA	Q6fly3 candida gia

32	161.5	9.1	479	2	Q91XE9_MOUSE	Q91xe9 m camp resp
33	161	9.1	585	2	Q4PBQ4_USTMA	Q4pbq4 ustilago ma
34	160.5	9.1	404	1	CRB3_MOUSE	Q61817 mus musculu
35	160	9.1	261	1	XBP1_HUMAN	P17861 homo sapien
36	159.5	9.0	807	2	Q51LL3_MAGGR	Q51ll3 magnaporthe
37	159	9.0	721	2	O22208_ARATH	O22208 arabidopsis
38	159	9.0	721	2	O8L7E7_ARATH	O8l7e7 arabidopsis
39	158.5	9.0	479	2	Q8BWS0_MOUSE	Q8bws0 mus musculu
40	157	8.9	616	2	Q524G5_MAGGR	Q524g5 magnaporthe
41	156	8.8	277	2	Q90513_FUGRU	Q90513 fugu rubrip
42	156	8.8	378	2	Q800A6_FUGRU	Q800a6 fugu rubrip
43	154.5	8.7	686	2	Q61D33_CAEBR	Q61d33 caenorhabdi
44	154.5	8.7	690	2	O44743_CAEBL	O44743 caenorhabdi
45	153.5	8.7	327	1	FOSL2_FAT	P51145 rattus norv

ALIGNMENTS

RESULT 1
Q8TFU8 EMENI PRELIMINARY; PRT; 350 AA.
AC Q8TFU8;
DT 01-JUN-2002 (Tremblrel. 21, Created)
DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)
DE 01-MAR-2004 (Tremblrel. 26, Last annotation update)
DE Transcription factor.
GN Name=hacA;
OS Emericella nidulans (Aspergillus nidulans).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; Emericella.
OX NCBI_TaxID=162425;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Saloheimo M.L.A., Valkonen M., Penttilae M.E.;
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL, AJ413273; CAC88375.1; -, Genomic_DNA.
DR HSSP; P05412; IJNM.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR011700; BZIP_2.
DR InterPro; IPR004827; TF_BZIP.
DR Pfam; PF07716; BZIP_2; 1.
DR SMART; SM00338; BRLZ; 1.
DR PROSITE; PS0217; BZIP; 1.
DR PROSITE; PS00036; BZIP_BASIC; UNKNOWN_1.
KW DNA-binding; Nuclear protein.
SQ SEQUENCE 350 AA; 38192 MW; 4F78340096064F44 CRC64;

Query Match 99.4%; Score 1755.5; DB 2; Length 350;
Best Local Similarity 99.7%; Pred. No. 7.5e-83;
Matches 349; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY	1	MKSADRFSPVKMEDAFANSP-PTTPSLEVPVLTVPADTSLRTKNVVAQTKPEKKPAKRR	59
DB	1	MKSADRFSPVKMEDAFANSLPTTPSLEVPVLTVPADTSLRTKNVVAQTKPEKKPAKRR	60
QY	60	KSWGQELVPKTNLPKRAKTEDEKQRRIRVLRNRAAQTSTRERKLEMEKLESEKI	119
DB	61	KSWGQELVPKTNLPKRAKTEDEKQRRIRVLRNRAAQTSTRERKLEMEKLESEKI	120
QY	120	DMEQNQFLLQRLAQWAEENRLSQVLAQSLAEVGRSHSTPTSSPASVPTLTPLFK	179
DB	121	DMEQNQFLLQRLAQWAEENRLSQVLAQSLAEVGRSHSTPTSSPASVPTLTPLFK	180
QY	180	QEGDEVPDRIPPPPTPSVTDYSPTLKPSLAESPDLTOHPAVSVGLEGDEGALTLDLG	239
DB	181	QEGDEVPDRIPPPPTPSVTDYSPTLKPSLAESPDLTOHPAVSVGLEGDEGALTLDLG	240
QY	240	ASIKHEPHTDLTAPLSDDDFRFLFNGDSLSLLEDGFAFDVLDSGDLSAFPFDSMW	299
DB	241	ASIKHEPHTDLTAPLSDDDFRFLFNGDSLSLLEDGFAFDVLDSGDLSAFPFDSMW	300

```
QY 300 DFDTEPVTLEDLEQTNGLSDSASCKAASLQPSHGASTSRCDGGIAAGSA 349
DB 301 DFDTEPVTLEDLEQTNGLSDSASCKAASLQPSHGASTSRCDGGIAAGSA 350

RESULT 2
Q5AQN3 EMENI PRELIMINARY; PRT; 347 AA.
AC Q5AQN3;
DT 10-MAY-2005 (TremBLrel. 30, Created)
DT 10-MAY-2005 (TremBLrel. 30, Last sequence update)
DE Hypothetical protein.
GN ORFName=AN9397.2;
OS Aspergillus nidulans FGSC A4.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eutotiales; Trichocomaceae; Emericella.
OX NCBI_TaxID=227321;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=FGSC A4;
RA Birren B., Nusbaum C., Abouelleil A., Allen N., Anderson S.,
RA Arachchi H.M., Barna N., Baetien V., Bloom T., Boguslavskiy L.,
RA Boukhgalter B., Butler J., Calvo S.E., Camarata J., Chang J.,
RA Choepl Y., Collymore A., Cook A., Cooke P., Corum B., DeArelano K.,
RA Diaz J.S., Dodge S., Dooley K., Dorris L., Elkins T., Engels R.,
RA Erickson J., Faro S., Ferreira P., Fitzgerald M., Gage D., Galagan J.,
RA Gardyna S., Gnerre S., Graham L., Grand-Pierre N., Hafez N.,
RA Hagopian D., Hagos B., Hall J., Horton L., Hulme W., Iliev I.,
RA Jaffe D., Johnson R., Jones C., Kamal M., Katat A., Karatas A.,
RA Kelle C., Landers T., Levine R., Lindblad-Toh K., Liu G., Lui A.,
RA Ma L.-J., Mabbitt R., MacLean C., Macdonald P., Major J., Manning J.,
RA Matthews C., Maucelli E., McCarthy M., Meldrum J., Meneus L.,
RA Mihova T., Mienga T., Murphy T., Naylor J., Nguyen C., Nicol R.,
RA Nielsen C.B., Norbu C., O'Connor T., O'Donnell P., O'Neil D.,
RA Oliver J., Peterson K., Phunkhang P., Pierre N., Purcell S.,
RA Rachupka A., Ramasamy U., Raymond C., Retta R., Riese C., Rogov P.,
RA Roman J., Schauer S., Schupack R., Seaman S., Severy P., Smirnov S.,
RA Smith C., Spencer B., Stange-Thomann N., Stojanovic N., Stubbs M.,
RA Talamas J., Tesfaye S., Theodore J., Topham K., Travers M.,
RA Vassiliev H., Venkataranan V.S., Viel R., Vo A., Wang S., Wilson B.,
RA Wu X., Wyman D., Young G., Zainoun J., Zembek L., Zimmer A., Zody M.,
RA Lander E.;
RT "Genome Sequence of Aspergillus nidulans";
RL Submitted (JAN-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; AACD01000172; EAA66464.1; -; Genomic_DNA.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
KW DNA-binding; Hypothetical protein; Nuclear protein.
SQ SEQUENCE 347 AA; 37841 MW; A7CAAFEE761B3E1 CRC64;

Query Match 96.0%; Score 1696; DB 2; Length 347;
Best Local Similarity 97.4%; Pred. No. 8.6e-80;
Matches 341; Conservative 1; Mismatches 4; Indels 4; Gaps 2;

QY 1 MKSADRFSPVKMEDAFANS-PTTPSLSEVPLTVSPADTSLRTKNVVA-QTKPEKKPAKPKR 59
DB 1 MKSADRFSPVKMEDAFANS-LTPPSLEVPLTVSPADTSLQTKNVVAQTKPEKKPAK 60
QY 60 KSWGQELPVKTNLPKRAKTEDEKEQRRIERVLNRNAAQTSRERKLEMEKLESEKI 119
DB 61 KSWGQELPVKTNLPKRAKTEDEKEQRRIERVLNRNAAQTSRERKLEMEKLESEKI 120
QY 120 DMEQNNQFLQRLAQMEANENRLSQVAQLSAEVRGSRHSTPTSSPASVSTLTPTLTK 179
DB 121 DMEQNNQFLQRLAQMEANENRLSQVAQLSAEVRGSRHSTPTSSPASVSTLTPTLTK 180
QY 180 QSGDEVPLDRIPFPPTSVTDYFTLPKPSLSAEPDLTQHPAVSVGGLGDESALTFLDLG 239
DB 181 QSGDEVPLDRIPFPPTSVTDYFTLPKPSLSAEPDLTQHPAVSVGGLGDESALTFLDLG 237

RESULT 4
Q4WEY8 ASPFU PRELIMINARY; PRT; 433 AA.
AC Q4WEY8;
DT 13-SEP-2005 (TremBLrel. 31, Created)
```

ID	Q8TFE3	TRIRE PRELIMINARY;	PRT;	451 AA.
AC	Q8TFE3;			
DT	01-JUN-2002	(TrEMBLrel. 21, Created)		
DT	01-JUN-2002	(TrEMBLrel. 21, Last sequence update)		
DT	01-MAR-2004	(TrEMBLrel. 26, Last annotation update)		
DE		Transcription factor.		
GN	Name=haci;			
OS	Trichoderma reesei (Hypocrea jecorina).			
OC	Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;			
OC	Hypocreomycetidae; Hypocreales; Hypocrea;			
OX	NCBI_TaxID=51453;			
RN	[1]			
RP	NUCLEOTIDE SEQUENCE.			
RA	Salohimo M.L.A., Valkonen M., Penttilae M.E.;			
RL	submitted (GSP-2001) to the EMBL/GenBank/DBJ databases.			
DR	EMBL; AJ131272; CAC88374.1; -; Genomic_DNA.			
DR	GO; GO:0005634; C:nucleus; IEA.			
DR	GO; GO:0003677; P:DNA binding; IEA.			
DR	GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.			
DR	InterPro; IPR011700; bZIP 2.			
DR	Pfam; PF07716; bZIP 2; I.			
DR	SMART; SM00338; BRLZ; 1.			
DR	PROSITE; PS0217; bZIP; 1.			
KW	DNA-binding; Nuclear protein.			
SQ	SEQUENCE	451 AA; 49277 MW; 600F10E471EA3AD3 CRC64;		
Query Match 24.6%; Score 434; DB 2; Length 451;				
Best Local Similarity 32.4%; Pred. No. 1.1e-14;				
Matches 149; Conservative 50; Mismatches 109; Indels 152; Gaps 21;				
Qy	5	DRFSPVMEHDAFANSPTTPSLEVPVLTVSPADTSLRTKNVAAQ-----		47
Db	26	DNFTSL-----FADS--TPS-----TLNPRD--MMTPDSVADISRLSVIPESQDAEDD		70
Qy	48	-----TKP--BEKPKAKRKWGQELVPVKTNLPPRKAAKTEDEKEQRRIERVLENRA		98
Db	71	ESHSTASATPSTSEKKPVKKRKGWQVLPKPKTNLPPRKAAKTEDEKEQRRIERVLENRR		130
Qy	99	AAOTSRERKLEMEKLESEK-----IDMEQOQNFLLQRLAQMEANNRLS-----		143
Db	131	AAQSRERKLEVEALEKRNKELETLINVKTNLILVEELNFRSSGVVTRSSSPDLS		190
Qy	144	-QQVAQLASAVRGSRH-----STPTSS-----SPASVSPTLTPTLFKQ----		180
Db	191	LQDSITLSQQLFGSRDQQTMSNPEQSLMDQIMRSAAANTVNPASLSPSLPPIIDKKEFQTK		250
Qy	181	EGDEVPLDIRPFPPTSPVSDYSPYLKPSSLAE-----SPDLTQHPAVSVGGLEGDESA		232
Db	251	EEDEEQADE---DEEMEQTWHETKEAATAKRNKSKQSRVSTDTQRPAVSIGG----DAA		303
Qy	233	LTLPF--DLGAS--IKHEPTHDLTAPLSDDDPRRLPFGSDSSLESDDLLEDGFADFVLDSGD		289
Db	304	VPVFSDDAGANGLGLDVPVHODGPFSS---IGHSEGLSAAILDADRYLLE-----SQ		350
Qy	290	LSAFPFDMSVDFD-----TEPVTLF-DLEQTNGLSDSASCKAASL-----		328
Db	351	LLASPNASTVDDYLAGDSAACTNPLPSDYDFDINDFLTDDANHAAYDIVAASNYAAAD		410
Qy	329	-----QPSHGASTSRCDGGQIAG		347
Db	411	RELDELTHDPENQTPSRHSTQQPSGASSHGCDGGIAGV		450
RESULT 6				
Q7SHF0_NEUCR				
ID	Q7SHF0	NEUCR PRELIMINARY;	PRT;	430 AA.
AC	Q7SHF0;			
DT	01-MAR-2004	(TrEMBLrel. 26, Created)		
DT	01-MAR-2004	(TrEMBLrel. 26, Last sequence update)		
DT	01-MAR-2004	(TrEMBLrel. 26, Last annotation update)		
DE		Predicted protein.		
GN	Name=NCU01856.1;			

```
OS Neurospora crassa.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.
OX NCBI_TaxID=5141;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=OR74A;
RA Galagan J.E., Calvo S.E., Borkovich K.A., Selker E.U., Read N.D.,
RA Jaffe D., FitzHugh W., Ma L.-J., Smirnov S., Purcell S., Reiman B.,
RA Elkins T., Engels R., Wang S., Nielsen C.B., Butler J., Endrizzi M.,
RA Qui D., Ianakiev P., Pedersen D., Nelson M., Washburne M.,
RA Sellitrennikoff C.P., Kinsey J.A., Braun E.L., Zelter A., Schulte U.,
RA Kothe G.O., Jedd G., Mewes W., Staben C., Marcotte E., Greenberg D.,
RA Roy A., Foley K., Naylor J., Thomann N., Barrett R., Gnerre S.,
RA Kamal M., Kamysyssel M., Mauceli E., Bielke C., Rudd S., Friseman D.,
RA Krystofova S., Ramussen C., Metzenberg R.L., Perkins D.D., Kroken S.,
RA Cogoni C., Macino G., Catchside D., Li W., Pratt R.J., Omani S.A.,
RA DeSouza C.C., Glass L., Orbach M.J., Berglund J., Voelker R.,
RA Yarden O., Plamann M., Seiler S., Dunlap J., Radford A., Aramayo R.,
RA Natvig D.O., Alex L.A., Mannhaupt G., Ebbole D.J., Freitag M.,
RA Paulsen I., Sachs M.S., Lander E.S., Nusbaum C., Birren B.,
RT "The Genome Sequence of the Filamentous Fungus Neurospora crassa.";
RL Nature 0:0-0(2003).
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; AABX01000004; EAA36251.1; -; Genomic_DNA.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR011700; bZIP 2.
DR Pfam; PF07716; bZIP 2; I.
DR PROSITE; PS50217; bZIP; 1.
SQ SEQUENCE 430 AA; 46599 MW; 0EB0657CFA6160DA CRC64;

Query Match 24.0%; Score 423; DB 2; Length 430;
Best Local Similarity 33.3%; Pred. No. 4e-14;
Matches 139; Conservative 35; Mismatches 137; Indels 106; Gaps 13;

QY 3 SADRFSPVKMED-----AFNSPTTSPLEVPVITVSPADTSLRTKNVVAQTKPEEKKP 55
DB 46 SLDVMSPVSLADDDVLTALATLTVPQIPPRSTPASTPAP-----ETEKKP 90

QY 56 AKKRKSWGQLPVPKTNLPKRAKTEDEKORRLERVLNRPAQAQTSREKRLEMEKLE 115
DB 91 VKKRKSWGQLPVPKTNLPKRAKTEDEKORRLERVLNRPAQAQTSREKRLEMEKLE 150

QY 116 SEKIDME-----QONQPLLQRLAQMEANNRLSQOVAQLSAEVRG----- 155
DB 151 RNNKELETLMQAQINQTLQALRE-----NGVAPTIAATPASPDLNPTPTVPSQELF 205

QY 156 -----SRH-----STPTSSSPASVSTPTTLTKQEGDEVPLDRIPFPPTS 196
DB 206 SSQDGHNLKSHDSLEQLPPTIKTBETVNPASLSPLNPLPEMEEDGKEQSTAAQPVAD 265

QY 197 VTDYSPTLKPSSLAESPDLTQHPAV-----SVGLEGGDSALTPLDLGASIKHEPTH 248
DB 266 AT-----STATVDTSPDQTPAVVFPVPAVSAPVAG-----SSISAAPLGLG---NEMD 311

QY 249 DLTAPLSDDDPRLRFGND--SSLESSSLEDGFAFDVLDSDGLSAPP---FDSMVDFDTE 304
DB 312 DLVLKSTFDADRYIYETEFSSPSFNFDEYSWAGDDTETLHFPANPNFFDEFINPDVS 371

QY 305 PVTLEDLQTNGLSDSASCAASL-----QPSHGASTSRCDGGOGIAAG 347
DB 372 NAANSNEQQPQQSGFCATTNVAHSVNLTYPNASEDPYQOPHTGASINGCDGGIAGV 428

RESULT 7
Q4HTT5_GIBZE
ID Q4HTT5_GIBZE PRELIMINARY; PRT; 429 AA.
AC Q4HTT5;
```



```

Db 344 IGDDNMVGDAPFNLNDDFISLWLNDDSAISAESWATSDFAAAIQGLEPKIYEPENQ-- 401
Qy 316 GLSDSASCKAASLQPSHGASTRCDCQGIAG 347
Db 402 -----VSSENPIQQPHPGASTQCDVGGIAG 428

RESULT 8
Q51KW8_MAGGR
ID Q51KW8_MAGGR PRELIMINARY; PRT; 556 AA.
AC Q51KW8;
DT 13-SEP-2005 (TrEMBLrel. 31, Created)
DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
DE Hypothetical protein.
GN ORFName=MG09010.4;
OS Magnaporthe grisea 70-15.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariomycetes; Incertae sedis; Magnaportheaceae; Magnaporthe.
OX NCBI_TaxID=242507;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=70-15;
RA Birten B., Nusbaum C., Abebe A., Abouelleil A., Adekoya E.,
RA Ait-zahra M., Allen N., Allen T., An P., Anderson M., Anderson S.,
RA Arachi H., Amruster J., Bachantsang P., Baldwin J., Barry A.,
RA Bayui T., Blishsteyn B., Bloom T., Blye J., Boguslavskiy L.,
RA Borsowsky M., Boukhgalter B., Brunahe A., Butler J., Calixte N.,
RA Calvo S., Camarata J., Campo K., Chang J., Cheshatsang Y., Citroen M.,
RA Collymore A., Consideine T., Cook A., Cooke P., Corum B., Cuomo C.,
RA David R., Dawoe T., Degray S., Dodge S., Dooley K., Dorje P.,
RA Dorjee K., Dorris L., Duffey N., Dupes A., Elkins T., Engels R.,
RA Erickson J., Farina A., Faro S., Ferreira P., Fischer H.,
RA Fitzgerald M., Foley K., Gage D., Galagan J., Gearin G., Gnerre S.,
RA Gierke A., Goyette A., Graham J., Grandbois E., Gyaltsen K., Hafez N.,
RA Hagopian D., Hagos B., Hall J., Hatcher B., Heller A., Higgins H.,
RA Honan T., Horn A., Houde N., Hughes L., Hulme W., Husby E., Iliev I.,
RA Jaffe D., Jones C., Kamal M., Kamat A., Kamysseilis M., Karlsson E.,
RA Kells C., Kieu A., Kisher P., Kodira C., Kulbokas E., Labutti K.,
RA Lana D., Landers T., Leger J., Levine S., Lewis D., Lewis T.,
RA Lindblad-toh K., Liu X., Lokyitsang T., Lokyitsang Y., Lucien O.,
RA Lui A., Ma L.J., Mabbitt R., Macdonald J., Maclean C., Major J.,
RA Manning J., Marabella R., Maru K., Matthews C., Mauceli E.,
RA McCarthy M., McDonough S., Mcghee T., Meldrim J., Meneus L.,
RA Mesirov J., Mihalev A., Mihova T., Mikkelson T., Mienga V., Moru K.,
RA Moses J., Mulrain L., Munson G., Naylor J., Neues C., Nguyen C.,
RA Nguyen N., Nguyen T., Nicol R., Nielsen C., Nizzari M., Norbu C.,
RA Norbu N., O'donnell P., Okoawo O., O'leary S., Omotosho B.,
RA O'Neill K., Oman S., Parker S., Perrin D., Phunkhang P., Pigani B.,
RA Purcell S., Rachupka T., Ramasamy U., Rameau R., Ray V., Raymond C.,
RA Retta R., Richardson S., Rise C., Rodriguez J., Rogers J., Rogov P.,
RA Rutman M., Schupbach R., Seaman C., Settipalli S., Sharpe T.,
RA Sheridan J., Sherpa N., Shi J., Smirnov S., Smith C., Sougnez C.,
RA Spencer B., Stalker J., Stange-thomann N., Stavropoulos S.,
RA Stetson K., Stone C., Stone S., Stubbs M., Talamas J., Tchuinga P.,
RA Tensing P., Tesfaye S., Theodore J., Thoultsang Y., Topham K.,
RA Towey S., Tsamla T., Tsomo N., Vallee D., Vassiliev H.,
RA Venkataraman V., Vinson J., Vo A., Wade C., Wang S., Wangchuk T.,
RA Wangdi T., Whittaker C., Wilkinson J., Wu Y., Wyman D., Yadav S.,
RA Yang S., Yang X., Yeager S., Yee E., Young G., Zainoun J., Zembeck L.,
RA Zimmer A., Zody M., Lander E.;
RT "The genome sequence of Magnaporthe grisea."
RL Submitted (OCT-2003) to the EMBL/GenBank/DBJ databases.
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=70-15;
RA Dean R., Mitchell T., Brown D., Pan H., Thon M.;
RL Submitted (OCT-2003) to the EMBL/GenBank/DBJ databases.
RN [3]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=70-15;
RA Zhu H., Blackmon B.;
RL Submitted (OCT-2003) to the EMBL/GenBank/DBJ databases.

```

```

CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; AACU01001606; EAA47880.1; -; Genomic_DNA.
DR InterPro; IPR011700; bZIP 2.
DR InterPro; IPR004827; TF_bZIP.
DR Pfam; PF07116; bZIP_2; 1.
DR SMART; SM00338; BRLZ; 1.
DR PROSITE; PS0217; bZIP; 1.
KW DNA-binding; Hypothetical protein; Nuclear protein.
SQ SEQUENCE 556 AA; 60798 MW; CFDC4C8D2DC4963C CRC64;

Query Match 19.6%; Score 346; DB 2; Length 556;
Best Local Similarity 36.3%; Pred.No. 4.9e-10;
Matches 103; Conservative 32; Mismatches 55; Indels 94; Gaps 12;

Qy 10 VKME----DAFANSP--TTPSL--EVPVLTVSPAD-----TSRLTKN---- 43
Db 10 LKWEASPAESFLSTPGDMYPSLFGDAKSFSLNPSDLASPPGSDPVELDAKTESAALR 69
Qy 44 -VVAQTKPEE-----KKPAKKGKSGQELVPKTNLPPRKRAKTEDEKQRRIVRL 94
Db 70 AVESPTPEETSSPAAASEKTKKKGSGQVLPKTNLPPRKRAKTEDEKQRRIVERVL 129
Qy 95 RNRAAAQTSRERKLEMEKLESEKIDMEQQOFLI--QLLAQMEANRNRLSQOVAQLSAR 152
Db 130 RNRAAQSRRERKQVEALEQRNALEQQ---LLHFQKLTQTMYOELQLRR----- 178
Qy 153 VRGSHSPTSSPASVPTTLTFLKQEGDEVPLDRIPFPFPPSV----- 197
Db 179 ----REAGVTSSEKPDGLTLPFLFRSQ-----DATASSATSVSATLADPANSLEDIF 228
Qy 198 --TVSPTLKPSLSAESP-----DLTQHPA 220
Db 229 RSTNVTTPNDPASIRSSPAPEQCHSVASGEAKTSADLTQHPA 272

RESULT 9
Q6CEV1_YARLI PRELIMINARY; PRT; 299 AA.
ID Q6CEV1_YARLI PRELIMINARY;
AC Q6CEV1;
DT 25-OCT-2004 (TrEMBLrel. 28, Created)
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE Yarrowia lipolytica chromosome B of strain CLIB99 of Yarrowia
DE lipolytica.
GN OrderedLocusNames=YALIOB12716g;
OS Yarrowia lipolytica (Candida lipolytica).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Dipodascaceae; Yarrowia.
OX NCBI_TaxID=4952;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=CLIB 122 / E 150;
RX PubMed=15229592; DOI=10.1038/nature02579;
RA Dujon B., Sherman D., Fischer G., Durens P., Casaregola S.,
RA Lafontaine I., de Montigny J., Marck C., Neuvéglise C., Talla E.,
RA Goffard N., Frangeul L., Aigle M., Anthouard V., Babour A., Barbe V.,
RA Barnay S., Blanchin S., Beckerich J.-M., Beyne E., Bleykasten C.,
RA Boerame A., Boyer J., Cattolico L., Confalieri F., de Daruvar A.,
RA Despons L., Fabre E., Fairhead C., Ferry-Dumazet H., Groppi A.,
RA Hantraye F., Hennequin C., Jauniaux N., Joyet P., Kachouri R.,
RA Kerrest A., Koszul R., Lemaire M., Lesur I., Ma L., Muller H.,
RA Nicaud J.-M., Nikolaki M., Oztas S., Ozier-Kalogeropoulos O.,
RA Pellenz S., Potier S., Richard G.-F., Straub M.-L., Suleau A.,
RA Swennen D., Tekalia F., Wesolowski-Louvel M., Westhof E., Wirth B.,
RA Zeniou-Meyer M., Zivanovic Y., Bolotin-Fukuhara M., Thierry A.,
RA Bouchier C., Caudron B., Scarpelli C., Gaillardin C., Weissenbach J.,
RT "Genome evolution in yeasts."
RL Nature 430:35-44(2004).
DR EMBL; CR382128; CAG83062.1; -; Genomic_DNA.
DR GO; GO:0005634; C:nucleus; IEA.

```


Qy

12 MEDAFANSPPTTSLPSLEVPVLTVSPADTSLRTKNVVAQTKEEKKPAKCKRSGWGLPVP-- 69

RN [5]

KL protein-response pathway.";
RL Genes Cells 1:803-817(1996).

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 23, 2005, 03:07:01 ; Search time 24.5265 Seconds
(without alignments)
1369.117 Million cell updates/sec

Title: US-10-663-450-4
Perfect score: 1766
Sequence: 1 MSAADRFSPVKMEDAFANSP.....PSHGASTSRCDGGIAGSA 349

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues
Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 80:.*
1: Pirl:.*
2: Pirl:.*
3: Pirl:.*
4: Pirl:.*

pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	231.5	13.1	238	2 S78571	transcription fact
2	165.5	9.4	322	2 T08592	TCAGC-motif-bindin
3	163.5	9.3	168	2 T50922	bZIP protein HV5 f
4	163	9.2	326	2 T08591	TCAGC-motif bindin
5	162	9.2	322	2 T12093	TCAGC-motif bindin
6	159.5	9.0	280	1 A36299	transcription fact
7	159	9.0	600	2 T00759	hypothetical prote
8	154.5	8.7	688	2 T32750	hypothetical prote
9	153.5	8.7	337	2 I55459	Fra-2 - rat
10	143.5	8.1	1165	2 T16420	hypothetical prote
11	142.5	8.1	380	1 TVRPF8	transforming prote
12	141	8.0	323	2 A35909	fra-2 protein - ch
13	139.5	7.9	2364	2 A56577	microtubule-associ
14	138.5	7.8	1737	2 A59235	unconventional myo
15	137	7.8	403	2 T47621	bZIP transcription
16	137	7.8	772	2 I50463	protein kinase - c
17	137	7.8	1001	2 T16419	hypothetical prote
18	136.5	7.7	326	2 I48351	fos-related antige
19	136.5	7.7	826	2 T43638	caspase-related pr
20	136	7.7	267	2 JC4857	hepatocarcinogenes
21	136	7.7	393	2 S16321	light-induced prot
22	134.5	7.6	380	1 TVMSFB	transforming prote
23	133	7.5	338	1 TVMSFB	transforming prote
24	133	7.5	381	1 TVMVJ	transforming prote
25	133	7.5	1820	2 T00365	hypothetical prote
26	133	7.5	1829	2 T24583	hypothetical prote
27	133	7.5	3938	2 T42761	Bassoon protein -
28	132.5	7.5	777	2 F54024	protein kinase (EC
29	132.5	7.5	1621	2 A82255	hypothetical prote

30	132	7.5	926	2 B37271	A-alpha Y 3 protei
31	131.5	7.4	380	1 TVHUF1	transforming prote
32	131.5	7.4	5327	2 T13564	microtubule-associ
33	130.5	7.4	326	2 S15749	transforming prote
34	130.5	7.4	392	2 E88108	protein C46E10.3 f
35	130	7.4	338	2 I53043	transforming prote
36	129.5	7.3	266	2 JC7300	tax-responsive ele
37	129.5	7.3	1255	2 T31065	diaphanous protein
38	129	7.3	3488	2 T34418	hypothetical prote
39	126	7.1	623	2 A48123	cell cycle regulat
40	125	7.1	313	2 A34785	DNA-binding protei
41	125	7.1	358	2 C42026	cyclic AMP respons
42	125	7.1	448	2 A42026	CAMP response elem
43	125	7.1	456	2 B42026	cyclic AMP respons
44	124.5	7.0	2464	1 QRMSPI	microtubule-associ
45	124.5	7.0	3942	2 T42730	Bassoon protein -

ALIGNMENTS

RESULT 1

S78571
transcription factor HAC1 - yeast (Saccharomyces cerevisiae)
N;Alternate names: protein YFL031w
C;Species: Saccharomyces cerevisiae
C;Date: 13-Sep-1998 #sequence_revision 13-Sep-1998 #text_change 09-Jul-2004
C;Accession: S78571; S56223; S53578
R;Murakami, Y.
submitted to the Protein Sequence Database, January 1998
A;Reference number: S78570
A;Accession: S78571
A;Molecule type: DNA
A;Residues: 1-238 <MUR>
A;Cross-references: UNIPROT:P41546; UNIPARC:UPI000157F61; EMBL:D50617; MIPS:YFL031w
A;Note: this is a revision to the sequence from reference S56186
R;Murakami, Y.; Naitou, M.; Hagiwara, H.; Shibata, T.; Ozawa, M.; Sasanuma, S.I.; Sasanuma
submitted to the EMBL Data Library, May 1995
A;Description: Analysis of the nucleotide sequence of chromosome VI from Saccharomyces cerevisiae
A;Reference number: S56186
A;Accession: S56223
A;Molecule type: DNA
A;Residues: 1-191, 'RWQTHSATI' <MUR>
A;Cross-references: UNIPARC:UPI000179B60; EMBL:D50617; NID:9836685; PID:9836723; MIPS:YFL031w
A;Note: this sequence has been revised in reference S78570
A;Note: this was believed to be the complete sequence of YFL031w
R;Nojima, H.; Leem, S.H.; Araki, H.; Sakai, A.; Nakashima, N.; Kanaoka, Y.; Ono, Y.
Nucleic Acids Res. 22, 5279-5288, 1994
A;Title: Hac1: a novel yeast bZIP protein binding to the CRE motif is a multicopy suppressor of the
A;Reference number: S53578; MUID:95116316; PMID:7816617
A;Accession: S53578
A;Molecule type: DNA
A;Residues: 1-142, 'RLCRPRCVRFRVGRDPFMAECLRRKMYQSRRLPYPTI', 183-220, 'AVITWTRKLQ' <NOJ>
A;Cross-references: UNIPARC:UPI000012C042; GB:D26506; NID:9633122; PID:BAA05513.1; PID:9633122
C;Genetics: SCD; HAC1; IRE2; ERN4
A;Gene: SCD; HAC1; IRE2; ERN4
A;Cross-references: SGD:S0001863; MIPS:YFL031w
A;Map position: 6L
A;Introns: 221/1
C;Keywords: DNA binding; nucleus; transcription factor

Query Match 13.1%; Score 231.5; DB 2; Length 238;
Best Local Similarity 29.0%; Pred. No. 5.6e-07;
Matches 79; Conservative 28; Mismatches 58; Indels 107; Gaps 10;

Qy	66	LPVP---KTNLPKRAKTEDEKQRIERVLRNRAAOTSRERKLEMEKLESEKIDME	122
Db	16	LAIPNFKSTLPKRAKTEDEKQRIERVLRNRAAOTSRERKLEMEKLESEKIDME	75
Qy	123	Q-QNOFLQRLQAE-----AENRLSQVAQLSAEVRGSRHSTPTSSSPAS	168
Db	76	NLNSVNLKLAHEDALTCSDHAFVASLDEYRDFQSTRGASLDTRASSHSSDSTPTSP	135

```
QY 169 VSPPTPTTLFKQGVDEVLDRIPFPPTPSVTDYSPTLKPSLSLAES----- 212
Db 136 LNCTMEPA-----TLSPKSMRDSASDQETSWELQMFKTE 169
QY 213 --PDLTQHAPVSVGGLEGDESALTLPDLGASTKHEPTHDLTAPLSDDDPRRFLPFGDSSLE 270
Db 170 NYPESTTLPAV-----DNN--NLFDVAVS-----PLADPLCDD-----IAGNS---- 205
QY 271 SDSSLLLEDGFADVLDSGLSAPFPDSDMYDFD 302
Db 206 -----LPPDNSIDL 215

RESULT 2
T08592
TGACG-motif-binding protein STF2 - soybean
C:Species: Glycine max (soybean)
C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 31-Dec-2004
C:Accession: T08592
R:Cheong, Y.H.; Yoo, C.M.; Park, J.M.; Ryu, G.R.; Goekjian, V.H.; Nagao, R.T.; Key, J.L.
submitted to the EMBL Data Library, September 1995
A:Description: STF1 is a novel TGACG-binding factor with a zinc-finger motif and a bZIP
A:Reference number: Z16445
A:Accession: T08592
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-326 <CHE>
A:Cross-references: UNIPROT:Q39896; UNIPARC:UPI00000A61C8; EMBL:L28004; NID:G986966; PID
A:Experimental source: strain Williams; hypocotyl
C:Gene: STF2
C:Superfamily: TGACG-motif-binding transcription factor

Query Match 9.4%; Score 165.5; DB 2; Length 322;
Best Local Similarity 27.8%; Pred. No. 0.0065;
Matches 47; Conservative 31; Mismatches 66; Indels 25; Gaps 2;

QY 9 PVKMDAFANSPPTPSLEVPLTVSPADTSLRTKNVVAQTKEPKPAKRRKSWGQLPV 68
Db 158 PCATSDAQSMQTTSIGQSEKVSLSVADPK-----QPGPESDEEIRRVPEIGESAG 209

QY 69 PKNLIP-----PRKRAKTEDEKORRIERVLRNRAAAQTKEPKPAKRRKSWGQLPV 111
Db 210 TSASQPDAGSNAGTERTVQGTGGQKGRSPADKESKRLKRLNRVSAQAARERKKAYL 269

QY 112 EKLESEKIDMEQONQFLQRLAQMENRLSQOVAQLSAEVRGSRHST 160
Db 270 IDLETRVKDLEKKNSSELKRLTLQENQMLRQILKNTTASRRGSNNGT 318

RESULT 3
T50922
bZIP protein HYS [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 21-Jul-2000 #sequence_revision 21-Jul-2000 #text_change 09-Jul-2004
C:Accession: T50922
R:Oyana, T.; Shimura, Y.; Okada, K.
submitted to the EMBL Data Library, July 1997
A:Description: The Arabidopsis HYS gene encodes a bZIP protein that regulates stimulus-i
A:Reference number: Z25271
A:Accession: T50922
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-168 <OYA>
A:Cross-references: UNIPROT:Q24646; UNIPARC:UPI000012CF04; EMBL:AB005295; PIDN:BA21116.
A:Experimental source: Landsberg erecta

Query Match 9.3%; Score 163.5; DB 2; Length 168;
Best Local Similarity 29.7%; Pred. No. 0.0039;
Matches 41; Conservative 28; Mismatches 68; Indels 1; Gaps 1;

QY 18 NSPTTSPSLEVPLTVSPADTSLRTKNVVAQTKEPKPAKRRKSWGQLPVPKTNLP 77
Db 18 NSPTTSPSLEVPLTVSPADTSLRTKNVVAQTKEPKPAKRRKSWGQLPVPKTNLP 77
```

```
Db 21 SSSAPHLEIKE-GIESDEEIRRVPEFGAEAVGKETSREGSGSATCQERTQATVGSQRK 79
QY 78 RAKTEDEKORRIERVLRNRAAAQTSSRKRLEMEKLESEKIDMEQONQFLQRLAQMBA 137
Db 80 RGRTPAEKRNKRLKRLNRVSAQAARERKKAYLSELENRVKDLNKNSELERLSTLQN 139
QY 138 ENNRLSQQOVAQLSAEVRG 155
Db 140 ENQMLRHILKNTTGNKRG 157

RESULT 4
T08591
TGACG-motif binding protein STF1 - soybean
C:Species: Glycine max (soybean)
C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 31-Dec-2004
C:Accession: T08591
R:Cheong, Y.H.; Yoo, C.M.; Park, J.M.; Ryu, G.R.; Goekjian, V.H.; Nagao, R.T.; Key, J.L.
submitted to the EMBL Data Library, September 1995
A:Description: STF1 is a novel TGACG-binding factor with a zinc-finger motif and a bZIP
A:Reference number: Z16445
A:Accession: T08591
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-326 <CHE>
A:Cross-references: UNIPROT:Q39895; UNIPARC:UPI00000A6233; EMBL:L28003; NID:G2934883; PID
A:Experimental source: strain Williams; hypocotyl
C:Superfamily: TGACG-motif-binding transcription factor

Query Match 9.2%; Score 163; DB 2; Length 326;
Best Local Similarity 30.4%; Pred. No. 0.0093;
Matches 45; Conservative 28; Mismatches 53; Indels 22; Gaps 3;

QY 20 PTPPSLE-----VPVLTVSPADTSLRTKNVVAQTKEPKPAKRRKSWGQLPVPKTN 72
Db 190 PKQPGPESDEEIRRVPEIGESAGTS-----ASRPDAGSNAGTERAQG-----RG 234

QY 73 LPPRKRAKTEDEKORRIERVLRNRAAAQTSSRKRLEMEKLESEKIDMEQONQFLQRL 132
Db 235 DSQKRGSPADKESKRLKRLNRVSAQAARERKKAYLIDLETRVKDLEKKNSSELKERL 294

QY 133 AQMEAENRLSQOVAQLSAEVRGSRHST 160
Db 295 STLQENQMLRQILKNTTASRRGSNGT 322

RESULT 5
T12093
TGACG-motif binding protein - fava bean
C:Species: Vicia faba (fava bean)
C:Date: 16-Jul-1999 #sequence_revision 16-Jul-1999 #text_change 31-Dec-2004
C:Accession: T12093
R:Wohlfarth, T.
submitted to the EMBL Data Library, May 1996
A:Reference number: Z17415
A:Accession: T12093
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-322 <WOH>
A:Cross-references: UNIPROT:O04234; UNIPARC:UPI00000A297E; EMBL:X97904
C:Superfamily: TGACG-motif-binding transcription factor

Query Match 9.2%; Score 162; DB 2; Length 322;
Best Local Similarity 29.7%; Pred. No. 0.01;
Matches 44; Conservative 27; Mismatches 55; Indels 22; Gaps 3;

QY 20 PTPPSLE-----VPVLTVSPADTSLRTKNVVAQTKEPKPAKRRKSWGQLPVPKTN 72
Db 186 PKQPGLESDEEMRRVPDGGESAGTSASHKGTGTAGPERAQTGEB---GQ----- 233

QY 73 LPPRKRAKTEDEKORRIERVLRNRAAAQTSSRKRLEMEKLESEKIDMEQONQFLQRL 132
Db 234 ---KKRGRSPADKESKRLKRLNRVSAQAARERKKAYLSDLETRVNDLEKKNSSELKERL 290
```


QY 133 AQMEAEENRLSQVQAQLSAEVRGRSHST 160
Db 291 STLQENQMLRQLKNTTASRGNGSGT 318

RESULT 6
A36299
transcription factor hXBP-1 - human
N/Alternate names: DNA-binding protein TREBS
C/Species: Homo sapiens (man)
C/Date: 10-Sep-1999 #sequence revision 10-Sep-1999 #text_change 09-Jul-2004
C/Accession: A36299, S12559, A47490
R/Liou, H.C.; Boothby, M.R.; Finn, P.W.; Davidson, R.; Nabavi, N.; Zeleznik-Le, N.J.; Tir
Science 247, 1581-1583, 1990
A/Title: A new member of the leucine zipper class of proteins that binds to the HLA DRAL
A/Reference number: A36299; MUID:90208323; PMID:2321018
A/Accession: A36299
A/Status: preliminary
A/Molecule type: mRNA
A/Residues: 1-260 <L10>
A/Cross-references: UNIPROT:P17861; UNIPARC:UPI0000031C60; GB:M31627; NID:g184485; PIDN:
R/Yoshimura, T.; Fujisawa, J.I.; Yoshida, M.
EMBO J. 9, 2537-2542, 1990
A/Title: Multiple cDNA clones encoding nuclear proteins that bind to the tax-dependent e
A/Reference number: S12559; MUID:90316112; PMID:2196176
A/Accession: S12559
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-32, 'GQA', 35-260 <YOS>
A/Cross-references: UNIPARC:UPI00006FC05; EMBL:X55543; NID:g287644; PIDN:CAA39149.1; PI
R/Ponath, P.D.; Fass, D.; Liou, H.C.; Glimcher, L.H.; Strominger, J.L.
J. Biol. Chem. 268, 17074-17082, 1993
A/Title: The regulatory gene, hXBP-1, and its target, HLA-DRA, utilize both common and d
A/Reference number: A47490; MUID:93352484; PMID:8349596
A/Accession: A47490
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-32, 'GQA', 35-128, 'T', 130-192, 'F', 194-198, 'R', 200-260 <PON>
A/Cross-references: UNIPARC:UPI000017330E; GB:L13850
A/Note: authors translated the codon ACC for residue 130 as Asn, AAG for residue 151 as
C/Genetics:
A/Gene: GDB:XBPL; XBP2
A/Cross-references: GDB:131393; OMIM:194355
A/Map position: 22pter-22qter
A/Intons: 76/3; 108/3; 151/3; 200/3
C/Superfamily: X box-binding protein 1; fos/jun DNA-binding domain homology
C/Keywords: DNA binding; leucine zipper; nucleus; transcription regulation
F;64-104/Domain: fos/jun DNA-binding domain homology <FJD>

Query Match 9.0%; Score 159.5; DB 1; Length 260;
Best Local Similarity 28.0%; Pred. No. 0.011;
Matches 55; Conservative 31; Mismatches 80; Indels 25; Gaps 6;

QY 15 AFANSTPTSLVPLTVSPADTSLRTKNVVAQTKEEKKPAKRKSWGQELVPKTNLP 74
Db 5 AAAPNPADGTPFKVLLSQGPAS-----AAGAPAAFLPMLVPAQRGASPEAASGGLP 55

QY 75 -PKRAK-TEDEKEQRIERVLNRAAAQTSRERKLEMEKLESEKIDMEQQOFLTL--- 129
Db 56 QAKRKQLTHLSPEEKALRKLNVAQATARDKKAWSLEQQVVDLEENQKLLLEN 115

QY 130 -----QRLAQMEAEENRLSQ-----VAQLSAEVRGRSHSTPTSSPASVSPILTPTLTK 179
Db 116 QLLREKTHGLVWENQELRGLMDALVAEEAEAKGN-EVRVPVAGSAESAALRLAPLQ 174

QY 180 QEGDEVPDLRI 190
Db 175 VQAQLSPLQNI 185

RESULT 7
T00759
hypothetical protein At2g40950 [imported] - Arabidopsis thaliana

N/Alternate names: hypothetical protein T20B5.15
C/Species: Arabidopsis thaliana (mouse-ear cress)
C/Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 31-Dec-2004
C/Accession: T00759; H84835
R/Rounsley, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; Kaul,
submitted to the EMBL Data Library, November 1997
A/Description: Arabidopsis thaliana chromosome II BAC T20B5 genomic sequence.
A/Reference number: Z14159
A/Accession: T00759
A/Status: translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-600 <ROU>
A/Cross-references: UNIPROT:O22208; UNIPARC:UPI0000179D30; EMBL:AC002409; NID:g2623294; I
A/Experimental source: cultivar Columbia
R/Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.; L
m.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.
Euse, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J
Nature 402, 761-768, 1999
A/Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A/Reference number: A84420; MUID:20083487; PMID:10617197
A/Accession: H84835
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-600 <STO>
A/Cross-references: UNIPARC:UPI0000179D30; GB:AE002093; NID:g2623309; PIDN:AAB86455.1; GS
C/Genetics:
A/Gene: T20B5.15; At2g40950
A/Map position: 2
C/Superfamily: bZIP transcription factor; fos/jun DNA-binding domain homology

Query Match 9.0%; Score 159; DB 2; Length 600;
Best Local Similarity 25.8%; Pred. No. 0.033;
Matches 70; Conservative 46; Mismatches 109; Indels 46; Gaps 9;

QY 5 DRFSVPKMEDAFANSTPTSLVPLTVL-----VSPA--DTSLRTKNVVAQTKEEK 53
Db 123 NRESPRDSDRCGADHNLDTPLSSQSGSCGSDSEATNNESSPKSRNVAVDQVKVE 182

QY 54 KPA-----KKRKSQWQELVPKTNLPKRKAKTED-----EKEQRIERVLNRAAA 100
Db 183 EAATTTTITKTKKEIDELTDESRSKYRRSGEDADASAVTGEDEKKRLMRNRESA 242

QY 101 QTSREKRLEMEKLESEKIDMEQQOFLQLQRLAQMEAEENRLSQVQAQLSAEVRGRSHST 160
Db 243 QLSRQRKHVYVELEEKVRNMHSTITDLNGKISYFAENATLRQ---QLGGNGMCPHLP 299

QY 161 PTSSSPASVSPILTPT-----LFKQEGDEVPDLRIPTFPPTSPVTDYSPTLKPSLLA 210
Db 300 P---PPMGYPMPAMPYPMPCPPYVVKQGSQVPLIPRLKPQNTLTGTSKAKSES 356

QY 211 ESDPLTQHPA-VSVGLEGDESALTFLDLGA 240
Db 357 KSEAKTKVASISFLGL-----LFCFLFLGA 382

RESULT 8
T32750
hypothetical protein F57B10.1 - Caenorhabditis elegans
C/Species: Caenorhabditis elegans
C/Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 09-Jul-2004
C/Accession: T32750
R/Greco, T.; Elliott, G.; Keppler, D.
submitted to the EMBL Data Library, December 1997
A/Description: The sequence of C. elegans cosmid F57B10.
A/Reference number: Z21219
A/Accession: T32750
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-688 <GRE>
A/Cross-references: UNIPROT:O44743; UNIPARC:UPI0000079D06; EMBL:AF039713; PIDN:AAB96719.1
A/Experimental source: strain Bristol N2; clone F57B10
C/Genetics:
A/Gene: CBSP:F57B10.1

A:Map position: 1
A:Introns: 54/1; 150/3; 196/1; 306/3; 383/3; 627/3

Query Match 8.7%; Score 154.5; DB 2; Length 688;
Best Local Similarity 25.7%; Pred. No. 0.073;
Matches 49; Conservative 41; Mismatches 74; Indels 27; Gaps 6;

QY 19 SPTTSPLEVPVLTVPADTSLRTKNVVAQTKEPKPAKKRSWGQELPVPKTNLP-- 76
DB 204 NPASISLNAPSSSFNPOSTS--STATSSSSSSSTNGGFKSTGERRKYFPLRLDESEI 261

QY 77 KRAKTE-----DEKEQRRIERVLNRNAAQTSRRKRKLEMEKLESEKIDMEQ 123
DB 262 KLCKKEGICLPDFFPLTKAEERDLKIRRKIRNKSQAQTSRRKKQDYIQELDRVSESTK 321

QY 124 QNOFLQRLAOMEAENRNLQQVAQLSAEV-RGSRHSTPTS-----SSPASVSPTLT 174
DB 322 ENQALKQQIERLSSENQSVISQLKLOAQLGQNAKRTTQAGRCLAVFMLSACLLVSPOLS 381

QY 175 PTLFKQEGDEV 185
DB 382 P-LGNQDNQKV 391

RESULT 9
Fsa-2 - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 02-Jul-1996 #sequence_revision 02-Jul-1996 #text_change 09-Jul-2004
C:Accession: I55459
R:Balser, R.D.; Klein, D.C.
J. Biol. Chem. 270, 27319-27325, 1995
A:Title: Circadian Expression of Transcription Factor Fra-2 in the Rat Pineal Gland.
A:Reference number: I55459; MUID:96070844; PMID:7592394
A:Accession: I55459
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-327 <RES>
A:Cross-references: UNIPROT:P51145; UNIPARC:UPI000012ABD0; EMBL:U18913; NID:g1001950; P1
C:Genetics:
A:Superfamily: fos transforming protein; fos/jun DNA-binding domain homology
F:120-160/Domain: fos/jun DNA-binding domain homology <FJD>

Query Match 8.7%; Score 153.5; DB 2; Length 327;
Best Local Similarity 24.2%; Pred. No. 0.034;
Matches 82; Conservative 43; Mismatches 121; Indels 93; Gaps 15;

QY 18 NSPTTSPLEVPVLTVPADTSLRTKNVVAQTKEPKPAKKRSWGQELPVP-----KT 71
DB 52 NAITTTSDQLQWVQPTVITSM--SNPYPRSHPSPLGLRVPVPGHMLPRPGVIKTIGT 109

QY 72 NLPFRKAK--TEDEKEQRRIERVLNRNAAQTSRRKRKLEMEKLESEKIDMEQNOFL 129
DB 110 TVGRRRRDQLSPDEEERIRRE-RNKLAARAKNRNRELTEKLTQTEELEEEKSGLQ 168

QY 130 ORLAOMEAENRNL-----QQVAQLSAEVGRGSRHSTPTS--SSPASVSPTLTPTLFKOE 181
DB 169 KGIAELQKEKEKLEFVLVAGHPCKISPE---ERRSPISGVQSLRGTSAGVPPVVKQE 225

QY 182 GDEVPLDRIPFPPTSPVDSPTLPKPSLAESPDLTQHPAVSVGGLEGDSALTFLDLGAS 241
DB 226 -----PPED-----SPSSA-GMDKQRSVIXPISAG-----GGF 256

QY 242 IKHEPTHDLTAPLSDDDFRRLFNGDSLSLSDSLLEDGFADFVLDGDLSPAPFDSMVD 301
DB 257 YGEEPLH---TFIVVTSTPAITPGTGNL-----VFTYPSV----- 288

QY 302 DTEPVTLEDTQNLGSLDSASCKAASLQPSHGASTSRCD 340
DB 289 -----LEQESPASPSESCKA-----HRRSSSGD 313

RESULT 10
Ti6420
hypothetical protein F52C9.8b - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 20-Sep-1999
C:Accession: Ti6420
R:Favella, T.
submitted to the EMBL Data Library, November 1995
A:Description: The sequence of C. elegans cosmid F52C9.
A:Reference number: Z18511
A:Accession: Ti6420
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1165 <FAV>
A:Cross-references: UNIPARC:UPI000017B9F5; EMBL:U39850; NID:g1055052; PID:g1055056; PIDN:
C:Genetics:
A:Gene: CESP:F52C9.8b
A:Introns: 21/1; 36/1; 61/3; 79/3; 107/3; 464/3; 566/1; 1010/3; 1029/3; 1106/3

Query Match 8.1%; Score 143.5; DB 2; Length 1165;
Best Local Similarity 23.1%; Pred. No. 0.62;
Matches 87; Conservative 57; Mismatches 140; Indels 93; Gaps 18;

QY 3 SADRFPVKMEDAFANSTPTTSPLEVPVLTVPADTSLRTKNVVAQTKEPKPAKKRSW 52
DB 528 SESKIEPVDVKPRVA--PVPP--QVPVTPTKPVITNNKKRIDVVVTLDED---APRRVQV 580

QY 63 GOELPVPKTNLPKRAKTEDEKEQRRIERVLNRNAAQTSR-ERKRLMEKLESE-KID 120
DB 581 KQEIPEVSTSDATKSDAAPTARGAVIRKQEVESDVAPNTILISAKKFERMKAEDKED 640

QY 121 ME-----QONQFLQRLAOMEA-----NNRLSQ-----QVAQLS-----AEVRGSR 157
DB 641 MKKKTAAALQEALFNIOERRRVEKEIAAFATTTNQAVPQNPQPASSVQIAQVSTSESAPGTS 700

QY 158 H--STPTSSSPASVSPTLTPTLPFKQEGDEVPLDRIPFP-----TPSVTDYSPTLKP 206
DB 701 EAAATETMTSPKTNVIVETEGEOEDE--DEIPIKKSKRRAKIVNSDDEEEVEVRHP 757

QY 207 SSLAESPDLTQHPAVSVGGLEGDESALTFLDLGASIK-----HEPTHDLTAPLSDDDF 259
DB 758 KRSDEKEKRH-----VSYAESDD-----DMPVVKKRRNQSPEDPEVSAASPSDEED- 806

QY 260 RLFLNGDSLSLSDSLLEDGFADFVLDGDLSPAPFDSMVDPTFPTVLTLEDLEQTNGL-- 317
DB 807 -----DDIGSFVVSNDNEDDA---DSFVVGDDSPIEVEEDEDDEDMIER 846

QY 318 -----SDSASCKAAS 327
DB 847 RSSRKRSDRSRKSAT 863

RESULT 11
TVRTPS
transforming protein fos - rat
C:Species: Rattus norvegicus (Norway rat)
C:Date: 31-Dec-1989 #sequence_revision 31-Dec-1989 #text_change 09-Jul-2004
C:Accession: A28263; I59159
R:Curran, T.; Gordon, M.B.; Rubino, K.L.; Sambucetti, L.C.
Oncogene 2, 79-84, 1987
A:Title: Isolation and characterization of the c-fos(rat) cDNA and analysis of post-transcriptional regulation of the c-fos gene.
A:Reference number: A28263; MUID:88143713; PMID:3325986
A:Accession: A28263
A:Molecule type: mRNA
A:Residues: 1-380 <CUR>
A:Cross-references: UNIPROT:P12841; UNIPARC:UPI000012AB6E; GB:X06769; NID:g55933; PIDN:CJ
R:Abate, C.; Luk, D.; Gentz, R.; Rauscher, F.O.
Proc. Natl. Acad. Sci. U.S.A. 87, 1032-1036, 1990
A:Title: Expression and purification of the leucine zipper and DNA-binding domains of Fos.
A:Reference number: I59159; MUID:90138931; PMID:2105492
A:Accession: I59159
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA

[illegible]

submitted to Genbank, September 1998
A;Description: Identification of two novel and highly divergent myosins in Dictyostelium
A;Reference number: A59235
A;Accession: A59235
A;Status: preliminary; not compared with conceptual translation
A;Molecule type: mRNA
A;Residues: 1-1737 <GEI>
A;Cross-references: UNIPROT:Q9TW28; UNIPARC:UPI000079FP9C; GB:AF090533; NID:G5714395; PIR:G5714395
A;Experimental source: strain AX2

Search completed: November 23, 2005, 03:26:04
Job time : 26.5265 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 23, 2005, 03:04:00 ; Search time 136.303 Seconds
(without alignments)
1125.017 Million cell updates/sec

Title: US-10-663-450-4
Perfect score: 1766
Sequence: 1 MKSADRFSPVKMEDAFANSP.....PSHGASTSRCDGGIAGSA 349

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_21.*
1: Geneseqp1980s.*
2: Geneseqp1990s.*
3: Geneseqp2000s.*
4: Geneseqp2001s.*
5: Geneseqp2002s.*
6: Geneseqp2003as.*
7: Geneseqp2003bs.*
8: Geneseqp2004s.*
9: Geneseqp2005s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1766	100.0	349	8	AdS12805 Aspergill
2	1766	100.0	349	8	AdS12803 Aspergill
3	1766	100.0	349	8	AdS12863 Aspergill
4	1755.5	99.4	350	4	AaB82976 Aspergill
5	1755.5	99.4	350	5	AaeI5372 Aspergill
6	1293.5	73.2	342	4	AaB82977 Aspergill
7	1293.5	73.2	342	5	AaeI5381 Aspergill
8	1293.5	73.2	342	5	AdS12818 Aspergill
9	1293.5	73.2	386	5	AaeI5379 Aspergill
10	1293.5	73.2	386	8	AdS12815 Aspergill
11	434	24.6	451	5	AaeI5371 Trichoder
12	434	24.6	451	8	AdS12801 Trichoder
13	434	24.6	451	8	AdS12804 Trichoder
14	419.5	23.8	450	4	AaB82975 Trichoder
15	398.5	22.6	409	7	ABO43144 A. thalia
16	398.5	22.6	409	7	AdB11925 Plant (A.
17	398.5	22.6	409	8	AdO2271 Thalesre
18	325	18.4	64	5	AaeI5374 Aspergill
19	288	16.3	64	5	AaeI5373 Trichoder
20	231.5	13.1	230	2	AaW53806 Transcrip
21	231.5	13.1	230	8	AdT87049 Yeast Str
22	231.5	13.1	238	2	AaW53807 Transcrip
23	216	12.2	200	8	AdS43437 Bacteri
24	192	10.9	376	7	Adf18647 Human spl

25	192	10.9	376	8	ADN16844	Adn16844 Human XB1
26	192	10.9	376	9	AD266508	Ad266508 Amino aci
27	192	10.9	376	9	AEb98685	Aeb98685 Endoplasm
28	192	10.9	928	9	ADY71599	Ady71599 Human IRE
29	178	10.1	84	5	ABP02534	Abp02534 Human ORF
30	176.5	10.0	68	5	AAeI5382	AAeI5382 Yeast HAC
31	176.5	10.0	68	8	ADs12859	AdS12859 Saccharom
32	164.5	9.3	634	8	ADP98908	Adp98908 C. albica
33	164	9.3	370	9	ADV86805	Adv86805 Transcrip
34	163.5	9.3	168	5	AAU93013	Aau93013 Arabidops
35	163.5	9.3	168	7	ADD30174	Add30174 Plant yie
36	163.5	9.3	168	8	ADi43893	Adi43893 Plant tra
37	163.5	9.3	211	3	AAG08861	Aag08861 Arabidops
38	160	9.1	261	6	ABr47631	AbR47631 Breast ca
39	160	9.1	261	7	ADb75629	AdB75629 Prostate
40	160	9.1	261	8	ADL83138	AdL83138 Human PRO
41	160	9.1	261	8	ADN16846	Adn16846 Human XB1
42	160	9.1	261	8	ADr14692	Adr14692 Human NF-
43	160	9.1	261	8	ABM82329	ABm82329 Tumour-as
44	160	9.1	261	9	ADX07539	AdX07539 Cyclin-de
45	160	9.1	261	9	ADY71598	Ady71598 Human non

ALIGNMENTS

RESULT 1
ADs12805
ID ADs12805 standard; protein; 349 AA.
XX AC ADs12805;
XX DT 16-DEC-2004 (first entry)
XX DE Aspergillus nidulans hacA DNA binding domain.
XX KW unfolded protein response; UPR; HAC1; PTC2;
XX KW unfolded protein response modulator; enzyme production; hacA; chaperone;
XX KW foldase; DNA binding domain.
XX OS Emericella nidulans.
XX PN US2004186070-A1.
XX PD 23-SEP-2004.
XX PF 15-SEP-2003; 2003US-00663450.
XX PR 24-MAR-2000; 2000US-00534692.
XX PR 23-MAR-2001; 2001US-00816277.
(GEMV) GENENCOR INT INC.
PI Penttila ME, Ward M, Wang H, Valkonen MJ, Saloheimo MLA;
XX WPI; 2004-707924/69.
Increasing the secretion of a heterologous protein, such as a therapeutic or an industrial enzyme, in genetically modified eukaryotic cells by inducing an elevated unfolded protein response (UPR).

Example 3; SEQ ID NO 6; 83pp; English.

The invention describes a method of increasing the secretion of a heterologous protein in a eukaryotic cell, comprising inducing an elevated unfolded protein response (UPR). Also described are: an isolated nucleic acid encoding a HAC1 protein, where the HAC1 protein induces unfolded protein response and has less than 50% similarity to yeast HAC1 protein; an isolated nucleic acid encoding a HAC1 protein, where the HAC1 protein induces unfolded protein response and where the HAC1 protein comprises a DNA binding region that has greater than 70% similarity to the DNA binding region of filamentous fungi HAC1 protein; a protein having unfolded protein response inducing activity and having greater

CC	enzymes. This is the amino acid sequence of <i>Aspergillus nidulans</i> HAC1									
CC	chaperone and foldase. Note: This sequence differs from ADS12803 given in									
CC	the sequence listing.									
XX										
QQ	Sequence 349 AA;									
	Query Match	100.0%;	Score 1766;	DB 8;	Length 349;					
	Best Local Similarity	100.0%;	Pred. No. 8.9e-140;							
	Matches 349;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;					
QY	1	MKSADRFSPVKMEDAFANSPPTPSLSEVPVLTVSPADTSLRTKNVVAQTKEPKKPAKRRK	60							
DB	1	MKSADRFSPVKMEDAFANSPPTPSLSEVPVLTVSPADTSLRTKNVVAQTKEPKKPAKRRK	60							
QY	61	SWGQELPVPKTNLPPRKRAKTEDEKQRIERVLRNRAAAQTSRRKRLEMEKLESEKID	120							
DB	61	SWGQELPVPKTNLPPRKRAKTEDEKQRIERVLRNRAAAQTSRRKRLEMEKLESEKID	120							
QY	121	MEQONQFLQRLAQMAEENRLSQVQAQIAEVRGSRHSTPTSSSPASVSPTLTPTLTKQ	180							
DB	121	MEQONQFLQRLAQMAEENRLSQVQAQIAEVRGSRHSTPTSSSPASVSPTLTPTLTKQ	180							
QY	181	EGDEVPLDRIPTPTSPVTDYSPTLKPSLAEPSDLTQHPAVSVGGLEGDESALTLDLGA	240							
DB	181	EGDEVPLDRIPTPTSPVTDYSPTLKPSLAEPSDLTQHPAVSVGGLEGDESALTLDLGA	240							
QY	241	SIKHEPTHDLTAPLSDDDFRRLFNGDSLSAESDSSLLLEDGFAPDVLDSGDLSAFPFDSMVD	300							
DB	241	SIKHEPTHDLTAPLSDDDFRRLFNGDSLSAESDSSLLLEDGFAPDVLDSGDLSAFPFDSMVD	300							
QY	301	FDTEPVTLEDQTNGLSDSASCKAASLOPSHGASTSRCDDGGQIAAGSA	349							
DB	301	FDTEPVTLEDQTNGLSDSASCKAASLOPSHGASTSRCDDGGQIAAGSA	349							
RESULT 4										
AAB82976	ID	AAB82976 standard; protein; 350 AA.								
XX	AC	AAB82976;								
XX	DT	11-SEP-2003 (revised)								
DT	DT	21-DEC-2001 (first entry)								
XX	XX	Aspergillus nidulans hacA, involved in unfolded protein response.								
DE	XX	HacA; transcription factor; unfolded protein response; protein secretion.								
XX	XX	Emericella nidulans.								
OS	XX	Key Location/Qualifiers								
FT	FT	Domain 53..116								
FT	FT	/label= DNA binding domain								
XX	XX	WO200172783-A2.								
PN	XX	04-OCT-2001.								
PD	XX	23-MAR-2001; 2001WO-US009401.								
PF	XX	24-MAR-2000; 2000US-00534692.								
XX	XX	(GEMV) GENENCOR INT INC.								
PA	XX	Penttila ME, Ward M, Wang H, Valkonen MJ, Saloheimo MLA;								
PI	XX	WPI; 2001-626252/72.								
DR	XX	N-PSDB; AAH26932.								
XX	XX	Increasing secretion of heterologous proteins e.g. lipase and cellulase								
PT	PT	in eukaryotic cells useful in industry to increase production and								
PT	PT	facilitate purification, by inducing an elevated unfolded protein								
PT	PT	response.								

XX	Claim 54; Fig 8A-B; 89pp; English.	PF	23-MAR-2001; 2001US-00816277.
PS		XX	
XX		PR	24-MAR-2000; 2000US-00534692.
CC	The present sequence is that of the hacA protein of Aspergillus nidulans,	XX	(GENV) GENENCOR INT INC.
CC	as deduced from the newly isolated hacA gene (see AAH26932). HacA protein	PA	
CC	is a transcription factor involved in the unfolded protein response	PI	Penttila ME, Ward M, Wang H, Valkonen MJ, Saloheimo MLA;
CC	(UPR). The invention provides methods for increasing the secretion of a	XX	
CC	heterologous protein in a cell by inducing an elevated UPR. This can be	PI	
CC	achieved by modulating the activity of HAC1 (or hacA), PTC2 or IRE1 in	XX	
CC	the cell, e.g. by gene overexpression. The cell from which the protein is	DR	WPI: 2002-033728/04.
CC	secreted can be any cell having an UPR, such as mammalian cells, insect	DR	N-PSDB; AAD24596.
CC	cells, yeast and filamentous fungi. The protein of interest can be any	XX	
CC	secreted protein such as a therapeutic protein or an industrial enzyme,	PT	Increasing secretion of heterologous proteins e.g. lipase and cellulase
CC	e.g. lipase, cellulase, endoglucanase-H, protease, carbohydrtase,	PT	in eukaryotic cells useful in industry to increase production and
CC	reductase, oxidase, isomerase, transferase, kinase, phosphatase, alpha-	PT	facilitate purification, by inducing an elevated unfolded protein
CC	amylase, glucoamylase, lignocellulose hemicellulase, pectinase and	XX	response.
CC	ligninase (claimed). (Updated on 11-SEP-2003 to standardise OS field)	PS	Claim 39; Fig 8; 56pp; English.
XX		XX	
SQ	Sequence 350 AA;	CC	The present invention relates to methods for increasing the secretion of
	Query Match 99.4%; Score 1755.5; DB 4; Length 350;	CC	heterologous protein in eukaryotic cells by inducing an elevated unfolded
	Best Local Similarity 99.7%; Pred. No. 6.8e-139;	CC	protein response (UPR). The method involves inducing the elevated UPR by
	Matches 349; Conservative 0; Mismatches 0; Indels 1; Gaps 1;	CC	increasing the presence of proteins such as HAC1, HACa, PTC2 or IRE1 in
		CC	cells. The method and sequences are useful for increasing the secretion
QY	1 MKSADRFSPVKMEDAFANS-PTTPSLEVPVLTVPADTSLRTKNVVAQTKPEKKPAKCR 59	CC	of heterologous proteins (e.g. lipase, cellulase, carbohydrtase) in
Db		CC	eukaryotic cells useful in industry to increase protein yields and to
	1 MKSADRFSPVKMEDAFANSLEPTTPSLEVPVLTVPADTSLRTKNVVAQTKPEKKPAKCR 60	CC	facilitate purification. The present sequence is Aspergillus nidulans
		CC	hacA protein. (Updated on 29-AUG-2003 to standardise OS field)
QY	60 KSWGQELVPKTNLPKRKAETDEKEQRRIRVLRNRAAAQTSRKRKLEMEKLESEKI 119	XX	
Db		SQ	Sequence 350 AA;
	61 KSWGQELVPKTNLPKRKAETDEKEQRRIRVLRNRAAAQTSRKRKLEMEKLESEKI 120		Query Match 99.4%; Score 1755.5; DB 5; Length 350;
QY	120 DMEQONQFLQLRQAQAEANNRLSQVAQLSAEVRGSRHSTPTSSPASVSPTLTPTLTK 179		Best Local Similarity 99.7%; Pred. No. 6.8e-139;
Db			Matches 349; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
	121 DMEQONQFLQLRQAQAEANNRLSQVAQLSAEVRGSRHSTPTSSPASVSPTLTPTLTK 180	QY	1 MKSADRFSPVKMEDAFANS-PTTPSLEVPVLTVPADTSLRTKNVVAQTKPEKKPAKCR 59
QY	180 QEGDEVPLDRIPFPPTPSVTDYSPTLKPSSLAESPDLTQHPAVSVGLEGDESALTFLDLG 239	Db	
Db			1 MKSADRFSPVKMEDAFANSLEPTTPSLEVPVLTVPADTSLRTKNVVAQTKPEKKPAKCR 60
	181 QEGDEVPLDRIPFPPTPSVTDYSPTLKPSSLAESPDLTQHPAVSVGLEGDESALTFLDLG 240	QY	60 KSWGQELVPKTNLPKRKAETDEKEQRRIRVLRNRAAAQTSRKRKLEMEKLESEKI 119
QY	240 ASIKEPHTDLTAPLSDDDFRFLFNGDSSLESDDSLLEDGFAFDVLDGDLGSAFFPDSMV 299	Db	
Db			61 KSWGQELVPKTNLPKRKAETDEKEQRRIRVLRNRAAAQTSRKRKLEMEKLESEKI 120
	241 ASIKEPHTDLTAPLSDDDFRFLFNGDSSLESDDSLLEDGFAFDVLDGDLGSAFFPDSMV 300	QY	120 DMEQONQFLQLRQAQAEANNRLSQVAQLSAEVRGSRHSTPTSSPASVSPTLTPTLTK 179
QY	300 DFDTEPVTLEDLEQTNGLSDSASCKAASLQPSHGASTSRCDCQGIAAGSA 349	Db	
Db			121 DMEQONQFLQLRQAQAEANNRLSQVAQLSAEVRGSRHSTPTSSPASVSPTLTPTLTK 180
	301 DFDTEPVTLEDLEQTNGLSDSASCKAASLQPSHGASTSRCDCQGIAAGSA 350	QY	180 QEGDEVPLDRIPFPPTPSVTDYSPTLKPSSLAESPDLTQHPAVSVGLEGDESALTFLDLG 239
RESULT 5		Db	
AAE15372			181 QEGDEVPLDRIPFPPTPSVTDYSPTLKPSSLAESPDLTQHPAVSVGLEGDESALTFLDLG 240
ID	AAE15372 standard; protein; 350 AA.	QY	240 ASIKEPHTDLTAPLSDDDFRFLFNGDSSLESDDSLLEDGFAFDVLDGDLGSAFFPDSMV 299
XX		Db	
AC	AAE15372;		241 ASIKEPHTDLTAPLSDDDFRFLFNGDSSLESDDSLLEDGFAFDVLDGDLGSAFFPDSMV 300
XX		QY	300 DFDTEPVTLEDLEQTNGLSDSASCKAASLQPSHGASTSRCDCQGIAAGSA 349
DT	29-AUG-2003 (revised)	Db	
DT	07-MAR-2002 (first entry)		301 DFDTEPVTLEDLEQTNGLSDSASCKAASLQPSHGASTSRCDCQGIAAGSA 350
XX		DE	Aspergillus nidulans hacA protein.
XX		XX	
KW	Heterologous protein secretion; unfolded protein response; UPR; lipase;	XX	
KW	cellulase; carbohydrtase; industry; purification; hacA protein.	XX	
XX		OS	Emericella nidulans.
XX		XX	
FT	Key Location/Qualifiers	XX	
FT	Binding-site 53..116	XX	
XX	/label= DNA-binding_domain	XX	
XX		PN	US2001034045-A1.
XX		XX	
PD	25-OCT-2001.	XX	
XX		XX	

304 EPTVLEDTQNGLSASCKAASLOPSHGASTSRCDGGGIAAG 347
 298 ESVGFEIGIEPHGLPDETSRQTSVQPSGLGASTSRCDGGGIAAG 341

Query Match 73.2%; Score 1293.5; DB 8; Length 342;
 Best Local Similarity 76.2%; Pred. No. 4.7e-100;
 Matches 262; Conservative 29; Mismatch 46; Indels 7; Gaps 3;

QY 5 DRFSVPMKEDAFANSPPTPSLEVPVLTVSPADTSRTRKVVVA-QTKPEKKPAKKKSG 63
 DB 4 EAFSPV---DSLPGSP---ELPLTVSPADTSRDDSSVQAGETKAEKKPVKKKSG 57

QY 64 QELPVKTNLPPRKAKTEDEKEORRIERVLNRNRAAQTSSRRKLEMEKLESEKIDMEQ 123
 DB 58 QELPVKTNLPPRKAKTEDEKEORRIERVLNRNRAAQTSSRRKLEMEKLESEKIDMEQ 117

QY 124 QNOFLQLRLAQMEAEENRLSQVAQLSAEVRGSRHSTPTSSSPASVPTLTPTLFKQEGD 183
 DB 118 QNOFLQLRLSQMEAEENRLNQVAQLSAEVRGSRGNTPKGSPVSPASPTLTPTLFKQERD 177

QY 184 EYPLDRIPPTPSVTDYSPTLKPSSLAESPDTQHPAVSVGGLEGDESALTFLDGLASIK 243
 DB 178 EYPLERIPPTPSITDYSPTLKPSSLAESPDTQHPAVSVGGLEGDESALTFLDGLASIK 237

QY 244 HEPTHDLTAPLSDDDPRFLFNGDSSLSLESDSLLLEGGFAFDVLDGGDLSAFPFDMSVDFDP 303
 DB 238 PHAADDLAAPLSDDDPRFLFNVDSVPGSSVLEGGFAFDVLDGGDLSAFPFDMSVDFDP 297

QY 304 EPTVLEDTQNGLSASCKAASLOPSHGASTSRCDGGGIAAG 347
 DB 298 ESVGFEIGIEPHGLPDETSRQTSVQPSGLGASTSRCDGGGIAAG 341

RESULT 9
 AAE15379
 ID AAE15379 standard; protein; 386 AA.
 AC AAE15379;
 XX
 XX 07-MAR-2002 (first entry)
 XX
 XX Aspergillus niger var. awamori haca protein #1.
 XX
 XX Heterologous protein secretion; unfolded protein response; UPR; lipase;
 KW cellulase; carbohydrase; industry; purification; haca protein.
 XX
 XX Aspergillus niger.
 OS
 XX US2001034045-A1.
 PN
 XX 25-OCT-2001.
 PD
 XX 23-MAR-2001; 2001US-00816277.
 PF
 XX 24-MAR-2000; 2000US-00534692.
 PR
 XX (GEMV) GENENCOR INT INC.
 PA
 XX Penttila ME, Ward M, Wang H, Valkonen MJ, Saloheimo MLA;
 PI
 XX WPI; 2002-033728/04.
 DR
 XX N-PSDB; AAD24601.
 DR
 XX Increasing secretion of heterologous proteins e.g. lipase and cellulase
 PT in eukaryotic cells useful in industry to increase production and
 PT facilitate purification, by inducing an elevated unfolded protein
 PT response.
 XX
 XX Claim 39; Fig 28; 56pp; English.
 PS
 XX The present invention relates to methods for increasing the secretion of
 CC heterologous protein in eukaryotic cells by inducing an elevated unfolded
 CC protein response (UPR). The method involves inducing the elevated UPR by
 CC increasing the presence of proteins such as HAC1, HACA, PTC2 or IRE1 in

304 EPTVLEDTQNGLSASCKAASLOPSHGASTSRCDGGGIAAG 347
 298 ESVGFEIGIEPHGLPDETSRQTSVQPSGLGASTSRCDGGGIAAG 341

Query Match 73.2%; Score 1293.5; DB 8; Length 342;
 Best Local Similarity 76.2%; Pred. No. 4.7e-100;
 Matches 262; Conservative 29; Mismatch 46; Indels 7; Gaps 3;

QY 5 DRFSVPMKEDAFANSPPTPSLEVPVLTVSPADTSRTRKVVVA-QTKPEKKPAKKKSG 63
 DB 4 EAFSPV---DSLPGSP---ELPLTVSPADTSRDDSSVQAGETKAEKKPVKKKSG 57

QY 64 QELPVKTNLPPRKAKTEDEKEORRIERVLNRNRAAQTSSRRKLEMEKLESEKIDMEQ 123
 DB 58 QELPVKTNLPPRKAKTEDEKEORRIERVLNRNRAAQTSSRRKLEMEKLESEKIDMEQ 117

QY 124 QNOFLQLRLAQMEAEENRLSQVAQLSAEVRGSRHSTPTSSSPASVPTLTPTLFKQEGD 183
 DB 118 QNOFLQLRLSQMEAEENRLNQVAQLSAEVRGSRGNTPKGSPVSPASPTLTPTLFKQERD 177

QY 184 EYPLDRIPPTPSVTDYSPTLKPSSLAESPDTQHPAVSVGGLEGDESALTFLDGLASIK 243
 DB 178 EYPLERIPPTPSITDYSPTLKPSSLAESPDTQHPAVSVGGLEGDESALTFLDGLASIK 237

QY 244 HEPTHDLTAPLSDDDPRFLFNGDSSLSLESDSLLLEGGFAFDVLDGGDLSAFPFDMSVDFDP 303
 DB 238 PHAADDLAAPLSDDDPRFLFNVDSVPGSSVLEGGFAFDVLDGGDLSAFPFDMSVDFDP 297

QY 304 EPTVLEDTQNGLSASCKAASLOPSHGASTSRCDGGGIAAG 347
 DB 298 ESVGFEIGIEPHGLPDETSRQTSVQPSGLGASTSRCDGGGIAAG 341

RESULT 9
 AAE15379
 ID AAE15379 standard; protein; 386 AA.
 AC AAE15379;
 XX
 XX 07-MAR-2002 (first entry)
 XX
 XX Aspergillus niger var. awamori haca protein #1.
 XX
 XX Heterologous protein secretion; unfolded protein response; UPR; lipase;
 KW cellulase; carbohydrase; industry; purification; haca protein.
 XX
 XX Aspergillus niger.
 OS
 XX US2001034045-A1.
 PN
 XX 25-OCT-2001.
 PD
 XX 23-MAR-2001; 2001US-00816277.
 PF
 XX 24-MAR-2000; 2000US-00534692.
 PR
 XX (GEMV) GENENCOR INT INC.
 PA
 XX Penttila ME, Ward M, Wang H, Valkonen MJ, Saloheimo MLA;
 PI
 XX WPI; 2002-033728/04.
 DR
 XX N-PSDB; AAD24601.
 DR
 XX Increasing secretion of heterologous proteins e.g. lipase and cellulase
 PT in eukaryotic cells useful in industry to increase production and
 PT facilitate purification, by inducing an elevated unfolded protein
 PT response.
 XX
 XX Claim 39; Fig 28; 56pp; English.
 PS
 XX The present invention relates to methods for increasing the secretion of
 CC heterologous protein in eukaryotic cells by inducing an elevated unfolded
 CC protein response (UPR). The method involves inducing the elevated UPR by
 CC increasing the presence of proteins such as HAC1, HACA, PTC2 or IRE1 in

CC cells. The method and sequences are useful for increasing the secretion
CC of heterologous proteins (e.g. lipase, cellulase, chaperone) in
CC eukaryotic cells useful in industry to increase protein yields and to
CC facilitate purification. The present sequence is *Aspergillus niger* var.
CC awamori haca protein
XX
SQ

Query Match 73.2%; Score 1293.5; DB 5; Length 386;
Best Local Similarity 76.2%; Pred. No. 5.6e-100;
Matches 262; Conservative 29; Mismatches 46; Indels 7; Gaps 3;
QY 5 DRFPVKMEDAFANSTPTTSLEVPVLTSPADTSLRTKNVVA-OTKPEKKPAKKKSWG 63
DB 48 EAFSPV---DSLGSPTP---ELPLTVSPADTSLDSSVQAGETKAEKKPVKKRKSNG 101
QY 64 QELPVKTNLPKRKRAKTEDEKEQRIERVLNRNAAQTSRERKLEMEKLESEKIDMEQ 123
DB 102 QELPVKTNLPKRKRAKTEDEKEQRIERVLNRNAAQTSRERKLEMEKLESEKIDMEQ 161
QY 124 QNFLLQRLSOMEAENNRNLSQVAQLSAEVRGSRHSTPTSSPASVSPPTLTPTLFKQEGD 183
DB 162 QNFLLQRLSOMEAENNRNLSQVAQLSAEVRGSRGNTPKPGSPVASPTLTPTLFKQERD 221
QY 184 EYPLDRIPPTPTSDYSPTLKPSLSAESPDLTQHPAVSVGGLEGDESALTFLDLGASTK 243
DB 222 EYPLDRIPPTPTSDYSPTLKPSLSAESPDLTQHPAVSVGGLEGDESALTFLDLGASTK 281
QY 244 HEPTHDLTAPLSDDDFRRLFNGLSDSLESDSLLDGFADFDVLDGDLGSAFFPDSMWDFDT 303
DB 282 PHAADDLAAPLSDDDFRRLFNGLSDSLESDSLLDGFADFDVLDGDLGSAFFPDSMWDFDP 341
QY 304 EYPTLEDLEQTNGLSDSASCKAASLOPSHGASTSRCDCGGIAAG 347
DB 342 ESVGFEIEPPHGLPDETSRQTSSVQPSLGASTSRCDCGGIAAG 385

RESULT 10
ADS12815
ID ADS12815 standard; protein; 386 AA.
XX
AC ADS12815;
XX
DT 16-DEC-2004 (first entry)
DE
XX Aspergillus nidulans haca chaperone and foldase #3.
KW unfolded protein response; UPR; HAC1; PTC2;
KW unfolded protein response modulator; enzyme production; haca; chaperone;
KW foldase; enzyme.
XX
OS Emericella nidulans.
XX
PN US2004186070-A1.
XX
PD 23-SEP-2004.
XX
PF 15-SEP-2003; 2003US-00663450.
XX
PR 24-MAR-2000; 2000US-00534692.
PR 23-MAR-2001; 2001US-00816277.
XX
XX (GENV) GENENCOR INT INC.
XX
PI Penttila ME, Ward M, Wang H, Valkonen MJ, Saloheimo MLA;
XX
DR WPI: 2004-707924/59.
DR N-PSDB; ADS12814.
XX
PT Increasing the secretion of a heterologous protein, such as a therapeutic
PT or an industrial enzyme, in genetically modified eukaryotic cells by
PT inducing an elevated unfolded protein response (UPR).
XX

PS Example 12; SEQ ID NO 16; 83pp; English.
XX
CC The invention describes a method of increasing the secretion of a
CC heterologous protein in a eukaryotic cell, comprising inducing an
CC elevated unfolded protein response (UPR). Also described are: an isolated
CC nucleic acid encoding a HAC1 protein, where the HAC1 protein induces
CC unfolded protein response and has less than 50% similarity to yeast HAC1
CC protein; an isolated nucleic acid encoding a HAC1 protein, where the HAC1
CC protein induces unfolded protein response and where the HAC1 protein
CC comprises a DNA binding region that has greater than 70% similarity to
CC the DNA binding region of filamentous fungi HAC1 protein; a protein
CC having unfolded protein response inducing activity and having greater
CC than 70% similarity to a fully defined amino acid sequence of 453, 349 or
CC 386, or 504 or 1332 amino acids (SEQ ID NO: 2, 4 or 16, or 12 or 14,
CC respectively) as given in the specification; a protein having an amino
CC acid sequence of SEQ ID NO: 2, 4, 8, 10, 12, 14 or 16; an isolated
CC nucleic acid encoding a PTC2 protein that modulates unfolded protein
CC response and has at least 70% similarity to a fully defined amino acid
CC sequence of 130 or 438 amino acids (SEQ ID NO: 8 or 10) as given in the
CC specification; an isolated nucleic acid encoding a IRE1 protein that
CC modulates unfolded protein response and has at least 60% similarity to a
CC fully defined amino acid sequence with SEQ ID NO: 12 or 14; a protein
CC having unfolded protein response modulating activity and having greater
CC than 70% similarity to SEQ ID NO: 8 or 10; and a cell containing a
CC heterologous nucleic acid encoding a protein having unfolded protein
CC response modulating activity and a heterologous nucleic acid encoding a
CC protein of interest to be secreted. The methods and compositions of
CC genetically manipulating cells to have an elevated unfolded protein
CC response (UPR) resulting in an increased capacity to produce secreted
CC proteins, are useful in e.g. production of therapeutic or industrial
CC enzymes. This is the amino acid sequence of *Aspergillus nidulans* HAC1
CC chaperone and foldase.
XX

SQ Sequence 386 AA;
Query Match 73.2%; Score 1293.5; DB 8; Length 386;
Best Local Similarity 76.2%; Pred. No. 5.6e-100;
Matches 262; Conservative 29; Mismatches 46; Indels 7; Gaps 3;
QY 5 DRFPVKMEDAFANSTPTTSLEVPVLTSPADTSLRTKNVVA-OTKPEKKPAKKRKSNG 63
DB 48 EAFSPV---DSLGSPTP---ELPLTVSPADTSLDSSVQAGETKAEKKPVKKRKSNG 101
QY 64 QELPVKTNLPKRKRAKTEDEKEQRIERVLNRNAAQTSRERKLEMEKLESEKIDMEQ 123
DB 102 QELPVKTNLPKRKRAKTEDEKEQRIERVLNRNAAQTSRERKLEMEKLESEKIDMEQ 161
QY 124 QNFLLQRLSOMEAENNRNLSQVAQLSAEVRGSRHSTPTSSPASVSPPTLTPTLFKQEGD 183
DB 162 QNFLLQRLSOMEAENNRNLSQVAQLSAEVRGSRGNTPKPGSPVASPTLTPTLFKQERD 221
QY 184 EYPLDRIPPTPTSDYSPTLKPSLSAESPDLTQHPAVSVGGLEGDESALTFLDLGASTK 243
DB 222 EYPLDRIPPTPTSDYSPTLKPSLSAESPDLTQHPAVSVGGLEGDESALTFLDLGASTK 281
QY 244 HEPTHDLTAPLSDDDFRRLFNGLSDSLESDSLLDGFADFDVLDGDLGSAFFPDSMWDFDT 303
DB 282 PHAADDLAAPLSDDDFRRLFNGLSDSLESDSLLDGFADFDVLDGDLGSAFFPDSMWDFDP 341
QY 304 EYPTLEDLEQTNGLSDSASCKAASLOPSHGASTSRCDCGGIAAG 347
DB 342 ESVGFEIEPPHGLPDETSRQTSSVQPSLGASTSRCDCGGIAAG 385
RESULT 11
AAE15371
ID AAE15371 standard; protein; 451 AA.
XX
AC AAE15371;
XX
DT 29-AUG-2003 (revised)
DT 07-MAR-2002 (first entry)
XX

DE Trichoderma reesei HAC1 protein.
KW Heterologous protein secretion; unfolded protein response; UPR; lipase;
KW cellulase; carbohydrase; industry; purification; HAC1 protein.
XX Hypocrea jecorina.
OS
FH Key Location/Qualifiers
FT Binding-site 84..147
FT /label= DNA-binding_domain
XX
PN US2001034045-A1.
XX
PD 25-OCT-2001.
XX
PF 23-MAR-2001; 2001US-00816277.
XX
PR 24-MAR-2000; 2000US-00534692.
XX
PA (GEMV) GENENCOR INT INC.
XX
PI Penttila ME, Ward M, Wang H, Valkonen MJ, Saloheimo MLA;
XX
DR WPI; 2002-033728/04.
DR N-PSDB; AAD24595.
XX
PT Increasing secretion of heterologous proteins e.g. lipase and cellulase
PT in eukaryotic cells useful in industry to increase production and
PT facilitate purification, by inducing an elevated unfolded protein
PT response.
XX
PS Claim 39; Fig 7; 56pp; English.
XX
CC The present invention relates to methods for increasing the secretion of
CC heterologous protein in eukaryotic cells by inducing an elevated unfolded
CC protein response (UPR). The method involves inducing the elevated UPR by
CC increasing the presence of proteins such as HAC1, HACA, PTC2 or IRE1 in
CC cells. The method and sequences are useful for increasing the secretion
CC of heterologous proteins (e.g. lipase, cellulase, carbohydrase) in
CC eukaryotic cells useful in industry to increase protein yields and to
CC facilitate purification. The present sequence is Trichoderma reesei HAC1
CC protein. (Updated on 29-AUG-2003 to standardise OS field)
XX
SQ Sequence 451 AA;
Query Match 24.6%; Score 434; DB 5; Length 451;
Best Local Similarity 32.4%; Pred No. 1.4e-27;
Matches 149; Conservative 50; Mismatches 109; Indels 152; Gaps 21;
QY 5 DRFSPVQMEDAFANGPTTPSLEVPVLTVPADTSLRTKNVVAQ----- 47
DB 26 DNFTSL-----FADS--TPS-----TLNPRD--NMTPDVADIDSRLSVIPESQDAEDD 70
QY 48 -----TKP--BEKKPAKRKSWGOELPVKTNLPKRKATEDEKQORIVRLNRA 98
DB 71 ESHSTSATAPSTSEKKPKVKRKSQGLVPKPTNLPKRKATEDEKQORIVRLNRR 130
QY 99 AAOQSRERKRLMEKLESEK-----IDMEQONQFLQRLQAQMEANNRLS----- 143
DB 131 AAOQSRERKRLVEALEKKNKELETLINQKTNLILVEELNFRSSGVVTRSSPLDS 190
QY 144 -QQAQLSAEVRGSRH-----STPTGS-----SPASVSPTLTPTLFKQ---- 180
DB 191 LQDSITLSQQLFGSRDQGTMSNPEQSLMDQIMRSAANPTVNPASLSPLSPISDFQTK 250
QY 181 EGDEVPDLRIPTPTSVTSTYTLKPSLAE-----SPDITQHPAVSVGLEGDESA 232
DB 251 EDEEQADE--DEEMEQTWHETKEAANAKEKNSKQSRVSTDTORPAVSIIG-----DAA 303
QY 233 LTLF--DLGAS--IKHEPTDLTAPLSDDDFRRLFGDSSLESDDSLLEDGFADFVLDSDG 289
DB 304 VPVFSDDAGANCLGLDPVHQDDGPFSS---IGHFGLSALDADRYLLE-----SQ 350

QY 290 LSAPFPDPMVDFD-----TEPVTL-DLEOTNGLSDSASCKAASL----- 328
DB 351 LLASPNASTVDDDDYLAGDSAACTNPLPSVDYDFDINFLTDDANHAAYDIVAASNYAAAD 410
QY 329 -----QPSHGASTSRCDGQGIAG 347
DB 411 RELDLEIHPENQIPSRHSIQQPSQGASSHGCDGDIAGV 450
RESULT 12
ADS12801
ID ADS12801 standard; protein; 451 AA.
XX
AC ADS12801;
XX
DT 16-DEC-2004 (first entry)
XX
DE Trichoderma reesei hac1 chaperone and foldase.
XX
KW unfolded protein response; UPR; HAC1; PTC2;
KW unfolded protein response modulator; enzyme production; hac1; chaperone;
KW foldase; enzyme.
XX
OS Hypocrea jecorina.
XX
PN US2004186070-A1.
XX
PD 23-SEP-2004.
XX
PF 15-SEP-2003; 2003US-00663450.
XX
PR 24-MAR-2000; 2000US-00534692.
PR 23-MAR-2001; 2001US-00816277.
XX
PA (GEMV) GENENCOR INT INC.
XX
PI Penttila ME, Ward M, Wang H, Valkonen MJ, Saloheimo MLA;
XX
DR WPI; 2004-707924/69.
DR N-PSDB; ADS12800.
XX
PT Increasing the secretion of a heterologous protein, such as a therapeutic
PT or an industrial enzyme, in genetically modified eukaryotic cells by
PT inducing an elevated unfolded protein response (UPR).
XX
Example 3; SEQ ID NO 2; 83pp; English.
The invention describes a method of increasing the secretion of a
heterologous protein in a eukaryotic cell, comprising inducing an
elevated unfolded protein response (UPR). Also described are: an isolated
nucleic acid encoding a HAC1 protein, where the HAC1 protein induces
unfolded protein response and has less than 50% similarity to yeast HAC1
protein; an isolated nucleic acid encoding a HAC1 protein, where the HAC1
protein induces unfolded protein response and where the HAC1 protein
comprises a DNA binding region that has greater than 70% similarity to
the DNA binding region of filamentous fungi HAC1 protein; a protein
having unfolded protein response inducing activity and having greater
than 70% similarity to a fully defined amino acid sequence of 451, 349 or
386, or 504 or 1232 amino acids (SEQ ID NO: 2, 4 or 16, or 12 or 14,
respectively) as given in the specification; a protein having an amino
acid sequence of SEQ ID NO: 2, 4, 8, 10, 12, 14 or 16; an isolated
nucleic acid encoding a PTC2 protein that modulates unfolded protein
response and has at least 70% similarity to a fully defined amino acid
sequence of 130 or 438 amino acids (SEQ ID NO: 8 or 10) as given in the
specification; an isolated nucleic acid encoding a IRE1 protein that
modulates unfolded protein response and has at least 60% similarity to a
fully defined amino acid sequence with SEQ ID NO: 12 or 14; a protein
having unfolded protein response modulating activity and having greater
than 70% similarity to SEQ ID NO: 8 or 10; and a cell containing a
heterologous nucleic acid encoding a protein having unfolded protein
response modulating activity and a heterologous nucleic acid encoding a
protein of interest to be secreted. The methods and compositions of
genetically manipulating cells to have an elevated unfolded protein

CC response (UPR) resulting in an increased capacity to produce secreted
CC proteins, are useful in e.g. production of therapeutic or industrial
CC enzymes. This is the amino acid sequence of *Trichoderma reesei* HAC1
CC chaperone and foldase.

Sequence 451 AA;

Query Match 24.6%; Score 434; DB 8; Length 451;

Best Local Similarity 32.4%; Pred. No. 1.4e-27;

Matches 149; Conservative 50; Mismatches 109; Indels 152; Gaps 21;

Qy 5 DRFSPVKMEDAFANSPPTTPSLEVPVLTVSPADTSLRTKNVVAQ----- 47

Db 26 DNFTSL-----FADS--TPS-----TLNPRD--MMTPDSVADIDSRLSVIPESQDAEDD 70

QY 48 -----TKP--EEKPAKKRKSQGOELPVKTNLPPRKRAKTEDEKEORRIERVLRNRA 98

Db 71 ESHSTSATPSTSEKPVKGRKSWGVLPEPKTNLPPRKRAKTEDEKEORRVERVLNR 130

99 AAOTSRERKRLEMEKLESEK-----IDMECONQFI.I.ORI.AQMEAFENNRI.S----- 143

```
db 131 AAOSSRRKRLEVEALEKRNKEIET.I.I.TNVOKTNI.I.IVEELNRRPRSSGVTRSSSP.DS 190
```

[illegible]

101	102	103	104	105	106	107	108	109	110	111	112	113	114	115	116	117	118	119	120	121	122	123	124	125	126	127	128	129	130	131	132	133	134	135	136	137	138	139	140	141	142	143	144	145	146	147	148	149	150	151	152	153	154	155	156	157	158	159	160	161	162	163	164	165	166	167	168	169	170	171	172	173	174	175	176	177	178	179	180	181	182	183	184	185	186	187	188	189	190	191	192	193	194	195	196	197	198	199	200
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100

[illegible][illegible][illegible][illegible]

Category	Sub-category	Value
A	1	..
	2	..
	3	..
	4	..
B	1	..
	2	..
	3	..
	4	..
C	1	..
	2	..
	3	..
	4	..
D	1	..
	2	..
	3	..
	4	..
E	1	..
	2	..
	3	..
	4	..
F	1	..
	2	..
	3	..
	4	..
G	1	..
	2	..
	3	..
	4	..
H	1	..
	2	..
	3	..
	4	..
I	1	..
	2	..
	3	..
	4	..
J	1	..
	2	..
	3	..
	4	..
K	1	..
	2	..
	3	..
	4	..
L	1	..
	2	..
	3	..
	4	..
M	1	..
	2	..
	3	..
	4	..
N	1	..
	2	..
	3	..
	4	..
O	1	..
	2	..
	3	..
	4	..
P	1	..
	2	..
	3	..
	4	..
Q	1	..
	2	..
	3	..
	4	..
R	1	..
	2	..
	3	..
	4	..
S	1	..
	2	..
	3	..
	4	..
T	1	..
	2	..
	3	..
	4	..
U	1	..
	2	..
	3	..
	4	..
V	1	..
	2	..
	3	..
	4	..
W	1	..
	2	..
	3	..
	4	..
X	1	..
	2	..
	3	..
	4	..
Y	1	..
	2	..
	3	..
	4	..
Z	1	..
	2	..
	3	..
	4	..

THE UNIVERSITY OF CHICAGO

DD 411 KEDDDEIHDFFENQIFSKHSIQFQSSC

RESULT 13

ADSL2804
TD ADSL2804 standard: protein: 451 aa

XX
AC
AND 13804.

XX
ED
16-DEC-2006 (F) 15:45:45

XX
E

[illegible]

KW unfolded protein response modulator; enzyme production; hac1; chaperone;

XX

XX
XX
XXXXXXX

Page 10

XX

XX
1002-350-07

XX
06-07-386-CT

PK 23-MAR-2001: 2001US-00816277.
 Z4-MAR-2000; 2000US-00334852.

XX
PA (GEMV) GENENCOR INT INC.

XX Penttilla ME Ward M Wan

QY 329 -----QPSHGASTSRCDGGIAAG 347

Db 411 RELDLEIHPENQIPSRHSIQQPSGASHCCDDGGIAGV 450

RESULT 14

AAB82975

ID AAB82975 standard; protein, 450 AA.

XX AAB82975;

XX 11-SEP-2003 (revised)

DT 21-DEC-2001 (first entry)

XX Trichoderma reesei HAC1, involved in unfolded protein response.

DE HAC1; transcription factor; unfolded protein response; protein secretion.

XX Hypocrea jecorina.

OS

XX

XX Key Location/Qualifiers

FT 84..147

FT Domain

FT /label= DNA binding domain

XX

XX WO200172783-A2.

XX

XX 04-OCT-2001.

XX

XX 23-MAR-2001; 2001WO-US009401.

XX

XX 24-MAR-2000; 2000US-00534692.

XX

XX (GENV) GENENCOR INT INC.

XX

XX Penttila ME, Ward M, Wang H, Valkonen MJ, Saloheimo MLA;

PI

XX WPI; 2001-626252/72.

DR N-PSDB; AAB26931.

XX

XX Increasing secretion of heterologous proteins e.g. lipase and cellulase

PT in eukaryotic cells useful in industry to increase production and

PT facilitate purification, by inducing an elevated unfolded protein

PT response.

XX

XX Claim 54; Fig 7A-B; 89pp; English.

PS

XX The present sequence is that of the HAC1 protein of Trichoderma reesei,

CC as deduced from the newly isolated HAC1 gene (see AAB26931). HAC1 protein

CC is a transcription factor involved in the unfolded protein response

CC (UPR). The invention provides methods for increasing the secretion of a

CC heterologous protein in a cell by inducing an elevated UPR. This can be

CC achieved by modulating the activity of HAC1, PTC2 or IRE1 in the cell,

CC e.g. by gene overexpression. The cell from which the protein is secreted

CC can be any cell having an UPR, such as mammalian cells, insect cells,

CC yeast and filamentous fungi. The protein of interest can be any secreted

CC protein such as a therapeutic protein or an industrial enzyme, e.g.

CC lipase, cellulase, endoglucanase-H, protease, carbohydratase, reductase,

CC oxidase, isomerase, transferase, kinase, phosphatase, alpha-amylase,

CC glucosylase, lignocellulose hemicellulase, pectinase and ligninase

CC (claimed). (Updated on 11-SEP-2003 to standardise OS field)

XX

XX Sequence 450 AA;

XX

Query Match 23.8%; Score 419.5; DB 4; Length 450;

Best Local Similarity 32.2%; Pred. No. 2.2e-26;

Matches 148; Conservative 50; Mismatches 109; Indels 153; Gaps 22;

QY 5 DRFPVKMEDANSPPTPSLEVPVLTVPATSLRTKNWAQ----- 47

Db 26 DNFTSL-----FADS--TFP-----TLNPRD--WMTSPSVADIDSRLSVIPESQDAEDD 70

QY 48 -----TKP--EEKPAKRRKSWGQELPVPKTNLPPRKRAKTEDEKEQRIERVLRNRA 98

Db 71 ESHSTATAPSTSEKPKVPKRWKSGVLPKPNLPPRKRAKTEDEKEQRIERVLRNRR 130

QY 99 AAOQTSRERKELEMEKLESEK-----IDMEQQNQFLQRLAQMEAEENRLS----- 143

Db 131 AAQSSRERKELEVEALEKRNKELETTLLINVQKTNLILVEELNRPFRSSGVVTRSSSPLDS 190

QY 144 -QVVAQLSAEVRGSRH-----STPTSS-----SPASVSTLTPTLFKQ----- 180

Db 191 LQDSITLSQQFLGSRDQGTMSNPESQSLMDQIMRSAAANPTVNPASLSPSLPISDKFQTK 250

QY 181 EGDEVPLDRIPFPPTPSVTDYSPTLKPSLAE-----SPDLTQHPAVSVGGLEGDESA 232

Db 251 EDEEQADE--DEMEQOTWHETKEAAAKKSKQSRVSTDTSTQFPA-SIGG-----DAA 302

QY 233 LTFLF--DLGAS-IKHEPTHTLTAPLSDDDFRRLPFGSDSLESLSLLEDGFAFVLDSDGD 289

Db 303 VPVFSDDAGNCLGLDPVHODDGPFS--ICHSGLSAALDADRYLLE-----SQ 349

QY 290 LSAPPPFDSMVDFF-----TEPVTLE-DLEQTNGLSASCKAASL----- 328

Db 350 LLASPNASTVDDDYLAGDSAACTNPLPSDYDFDINDFLTDDANHAAYDIVAASNYAAAD 409

QY 329 -----QPSHGASTSRCDGGIAAG 347

Db 410 RELDLEIHPENQIPSRHSIQQPSGASHCCDDGGIAGV 449

RESULT 15

AAB43144

ID ABO43144 standard; protein, 409 AA.

XX ABO43144;

XX 23-SEP-2003 (first entry)

XX A. thaliana disease tolerance transcription factor, G1034.

DE Plant; transcription factor; disease resistance; transgenic;

XX plant breeding; pathogens resistance; pests; resistance.

XX Arabidopsis thaliana.

OS

XX US2003046723-A1.

XX

XX 06-MAR-2003.

XX

XX 22-MAR-2000; 2000US-00533029.

XX

XX 22-MAR-2000; 2000US-00533029.

XX (HEAR/) HEARD J.

PA (BROU/) BROUN P.

PA (RIEC/) RIECHMANN J L.

PA (KEDD/) KEDDIE J.

PA (PINE/) PINEDA O.

PA (ADAM/) ADAM L.

PA (SAMA/) SAMAHA R.

PA (ZHAN/) ZHANG J.

PA (YUGG/) YU G.

PA (RATC/) RATCLIFFE O.

PA (PIUG/) PILGRIM M.

PA (JIAN/) JIANG C.

PA (REUB/) REUBER L.

XX

XX Heard J, Broun P, Riechmann JL, Keddie J, Pineda O, Adam L;

PI Samaha R, Zhang J, Yu G, Ratcliffe O, Pilgrim M, Jiang C, Reuber L;

XX WPI; 2003-521768/49.

DR N-PSDB; ACD98410.

XX

XX New transgenic plants comprising a recombinant gene that alters the

PT plant's disease tolerance or resistance, useful in plant breeding, e.g.

PT for generating plants with improved tolerance or resistance to diseases,

THIS PAGE BLANK (USPTO)


```
QY 177 SSGVTRSSPLDSLQDS-----ITLSQQLFGSRDGTMSNPESQLMDQIMRGAANPTVNP 232
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 169 TNPLSYSDSTPDISDSQSLPSTFKQLFNAQD--ELCRP---ISPQSIGPLTSRTVDP 223
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 233 ASLSPSLPISDKFQTKEDBEQADEDEMEQTHETKEAAAAKNSKQSRVSTDSIQ 292
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 224 STLSP-----KSLSPDSSNSN-----SSDMTQ 246
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 293 RPAV 296
  |||
Db 247 HPAV 250

RESULT 2
US-09-248-796A-18840
; Sequence 18840, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 18840
; LENGTH: 248
; TYPE: PRT
; ORGANISM: Candida albicans
; NAME/KEY: UNSURE
; LOCATION: (221)
; OTHER INFORMATION: Identity of amino acid sequences at the above locations are unkno
US-09-248-796A-18840

Query Match 11.3%; Score 258.5; DB 2; Length 248;
Best Local Similarity 29.7%; Pred. No. 4.6e-14;
Matches 80; Conservative 43; Mismatches 65; Indels 81; Gaps 10;

QY 31 LFADSTPSTLNPRDM--MTPDVSADIDSRLSVIPESQDAEDDESHSTSATSTSEKPV 88
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 21 LTVDTNTTNSDDLSVAIPTSLM-----TSITTSFMS----- 54
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 89 KKRKSGQVLP-EP---KTNLPPRKAKTEDEKQRRVERVLNRRAAQSSRRKRLEVE 144
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 55 TSTSSHNTLDDIDPATFKSTLPFRKAKTQEEKEQRIILNRRAAHASREKKRKHYE 114
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 145 ALEKRNKELETLINVOKTNLILVEELNFRSSGVVTRSS-----SPLDSIQDSITLSQ 199
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 115 YLENYVLKLET---NUMKLN-----NNYNOAFELLTKDNQELGLSKLEVLDDVSDLKE 164
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 200 QLFGRDGTGTMNPEOSLMDQIMRGAANPTVNPASLSPISDKFQTKEDBEQADE 259
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 165 QIHSNWSGTRSHNKS-----NDEDEEEDDE 191
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 260 DEEMEQTHETKEAAAAKNSKQSRVST 288
  ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 192 DEHLEE-GHVEQKEIKKEEPVSKRKLTNT 219
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 3
US-09-248-796A-14747
; Sequence 14747, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; FILE REFERENCE: 107196.132
```

```
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 14747
; LENGTH: 901
; TYPE: PRT
; ORGANISM: Candida albicans
; US-09-248-796A-14747

Query Match 6.5%; Score 149.5; DB 2; Length 901;
Best Local Similarity 21.7%; Pred. No. 0.0061;
Matches 78; Conservative 59; Mismatches 142; Indels 81; Gaps 15;

QY 30 SLFADSTPSTLNPRDMTPDVSADIDSRLSVIPESQDAEDDESH-----STSATA 79
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 455 ALISDDPYSTIEQLDITMVEEVLHGIDNTLTLDTKSDTKSEKPFVAEPKSVDTKKELE 514
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 80 PSTSEKPKVKKESWG-QVLPEPKT--NLPPRKAKTED-----EKEQRRVERVLNRN- 128
  |||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 515 TKTSEKS--EESKELGVKLEKVDTEKEIPVEESKKVDDNDDETDLKESKETKVEKASDKP 572
  |||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 129 -----RRAAQSSRRERKRLEVEALEKRNKELETLINNVQKTNLILVEELNFRRS 177
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 573 ITHEELIDEDIEKAKHQSNKDDANDTVLAE-----EGALVNEDDYDLVDKKELEK- 623
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 178 SGVTRSSPLD--SLQDSITLSQQLFGSRDGTWNSNPESQLMDQIMRGAANPTVNPASL 235
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 624 EGITTEEAPLNIEKKSEKTKVEDVLESSVKDDKPKILDVESKEAKKAADDEKAASE 683
  : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 236 SPSLPPI--SKEFQTKEDE-----EQADED-----EEMEQTWH----- 268
  ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 684 QPPLDIVKSKSEKTEKEDTTTKSELPGTKVEDSEKDTTTLKSEVESELEKSEEQPLDI 743
  ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 269 -----ETKEAAAAKNSKQSRVS--TDSTORPAVSIGDAAVPVFSDDAGANCLGLDP 320
  ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 744 KKEVVEYTKDDVATEKSKDKDVEQAVSSTTKETTKPEV-LETEKPKPAVADDDDLDDLDISP 802
  ||| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 4
US-09-487-558B-430
; Sequence 430, Application US/09487558B
; Patent No. 6949356
; GENERAL INFORMATION:
; APPLICANT: Busby, Robert
; APPLICANT: Cali, Brian
; APPLICANT: Hecht, Peter
; APPLICANT: Holtzman, Doug
; APPLICANT: Madden, Kevin
; APPLICANT: Maxon, Mary
; APPLICANT: Milne, Todd
; APPLICANT: Royer, John
; APPLICANT: Salama, Sofie
; APPLICANT: Sherman, Amir
; APPLICANT: Silva, Jeff
; APPLICANT: Summers, Eric
; TITLE OF INVENTION: Methods
; FILE REFERENCE: 109272.130
; CURRENT APPLICATION NUMBER: US/09/487,558B
; CURRENT FILING DATE: 2000-01-19
; PRIOR APPLICATION NUMBER: US 60/487,558
; PRIOR FILING DATE: 1999-10-20
; NUMBER OF SEQ ID NOS: 446
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 430
; LENGTH: 650
; TYPE: PRT
; ORGANISM: Saccharomyces cerevisiae
; US-09-487-558B-430
```

```

US-09-949-016-6835

Query Match      6.4%; Score 147.5; DB 2; Length 2058;
Best Local Similarity 23.5%; Pred. No. 0.003;
Matches 86; Conservative 58; Mismatches 119; Indels 103; Gaps 18;

QY      91  RKSQGVLPKPKTNLP RK-----RAKTEDEKEQRRVERVLNRNRAAOSSRRKRKLEVEA 145
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      805  RRVTRQLAEKREGEKKQEEEEKKKREERERERRERAEALRAQBEETRKKQEELEA 864
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY      146  LEKENKELE-TLLINQKTNLILVEELNFRFRSGVWTRSSPLDSQDSITLSQQLFGS 204
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      865  LQKSQKEAELTRELEKQENK-QVEEILRLRK-----EIEDLQRMKEQ----- 906
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY      205  RDGQTMNPGSLMDQIMRSAAANTVPASLSPSLPISDKFQTKKEDESEA----- 257
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      907  ---QELSLTEASQ-----KQERRDQELRLREAEACRAAQEFLES 944
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY      258  ---DEDEE---MBQTHWETKEAAAAKEKNSKQSRVSTDSQ---RPAVSIQGGDAAPVF 307
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      945  LNFDEIDECVRNIERSLSVGEFSSSELAESAECBKPNFQPYPEEVEDEGEFADDDAF 1004
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY      308  SDDAGANCLGLDPVHQDDGPFISGHS-----RGLSAAALDA---DRYLLESQLLASNA-ST 359
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      1005  KD-----SPNFSEHGHSQRTSGIRTSDDSSSEDPYMNQTVPTSPSADST 1050
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY      360  V-----DDLYL---AGDSAACF---TNPPLPS---DYDFDINDFLTDDANHAAYDIVA 402
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      1051  VLLAPSVQDSGLNSHSSGSESTYCMNPQVAGLPSPDGDDYDYDDQDY-EDGAI TSGSSVTF 1109
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
QY      403  ASNYAA 408
      | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db      1110  SNSYGS 1115

RESULT 6
US-09-949-016-10199
; Sequence 10199, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 10199
; LENGTH: 2111
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-10199

```

Query Match	6.4%	Score 147.5;	DB 2;	Length 2111;
Best Local Similarity	23.5%;	Pred. No. 0.0031;		
Matches	86;	Conservative	59;	Mismatches 119; Indels 103; Gaps 18;
Qy .	91	RKSWGVLPEPKTNLPKR-----RAKTEDEKQRVVRVLRNRAQAOSRERKKEVEEA	145	
		: : : : : :	:	:
Ddb	858	RRYVROLLAKREQEKKQEEBEEKKREEEREREREAEALRAQQEBETRKQOELEA	917	
		: : : : : :	:	:
Qy	146	LEYRNKELE--TLINIVQKNLILVLEELNRPRSSGVVTRSSSLDSDITILSQQLFGS	204	
		: : : : : :	:	:
Ddb	918	LQKSQKAELURELEKQENK-QVEILRLUEK-----ETEDLQRNKEQ----	959	
		: : : : : :	:	:
Qy	205	RDGQTMSNPQSIMDDGIMRSAAANPTVPASLSPSLPPISDKFQTKDEEDEQA-----	257	
		: : : : : :	:	:

Db 960 ---QELSLTEASLQ-----KLQERRDQBLRLREAEACRAAQEFLES 997
QY 258 ---DEDEE---MEQTHWETKEAAAKENKQSRVSTDTSTQ---RPAVISIGDDAAVPVF 307
Db 998 LNFDEIDECVRIERSLSVGSFELASAECEKPNFNSQFPYEEVEDEGEFADDDAF 1057
QY 308 SDDAGANCLGLDPVHQDDGPPFGHGS-----FGLSAAALDA---DRYLLESQLLASPNA-ST 359
Db 1058 KD-----SPNPSEHGSHSDQRTSGIRTSDDSEEDPYWNDIVVTSPSAUST 1103
QY 360 V-----DDDYL-----AGDSAACF---TNPLPS---DYDFDINDFLTDANHAAYDIVA 402
Db 1104 VLLAPSVQDSGLHNSSGESTYCMQPNAGDLPSPDGDYDQDDY-EDGALTSGSVTF 1162
QY 403 ASNYAA 408
Db 1163 SNSYGS 1168
RESULT 7
US-09-248-796A-14492
; Sequence 14492, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; PRIOR FILING DATE: 1999-02-12
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 14492
; LENGTH: 667
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-248-796A-14492
Query Match 6.4%; Score 146.5; DB 2; Length 667;
Best Local Similarity 23.6%; Pred. No. 0.00071;
Matches 87; Conservative 62; Mismatches 146; Indels 73; Gaps 16;
QY 56 SRLSVIPESQDAEDDESHSTATAPSTSEK-KPVKKRKSQGVLPPEKTNLPPEK----- 109
Db 139 SAVTASSASATPEPENESLAASEQNEAPKAPKAKKPKAGLAAALKKQLELKKQLEEE 198
QY 110 RAKTEDEKEORV-----ERVLNRRAAQSSRRERKLEVEALE-----KRKNELET 155
Db 199 QRRLEEEERKLEERLAEEBAEAKKAAAKKRRERLUKQKQKQKQKQKQKQKQKQKQKQK 258
QY 156 L-----LI-----NYQKTNLILVEELNRRPRSSGVVT--RSSPSLDSLQDSITLSQQLFG 203
Db 259 LQERRRQQLQAGNSVAGLSKTEGDAPKPKVYVTKKSKTKPTFIQKTPV-SKAPVK 317
QY 204 SRDQTMNPSQSLMDQIMRSANPTVNDPA-SLSPSLPPISDKFQTKKEDEEQADEDEE 262
Db 318 KDDGE-----DEALVDDMEKMLDDBAPVADPWAEALADEGEERVEGEAEDEEVEGEDEID 372
QY 263 -----MEQTHWETKEAAA---AKEKNSQSRVSTDTSTORPAVISIGGDAAPVPSD 309
Db 373 TIVEDEQEKKAEEAARKAEEAARAKAEEAARKAEEAARKAEEAARKAEEAARKAEEAARK 429
QY 310 DAGANCLGLDPVHQDDGPPFSI-----GHSFGLSAAALDA-----DRYLLESQLLA 353
Db 430 LRSPICCILG--HVDTGKTKLDKIRQTNVQGEAGGIQIQIGATVFPVDAIKQKTAVMA 487
QY 354 SPNASTVD 361
Db 488 KYEKQTFD 495

RESULT 8
US-10-012-231A-308
; Sequence 308, Application US/10012231A
; Patent No. 6924355
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan I.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth J.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; FILE REFERENCE: P2830PIC23
; CURRENT APPLICATION NUMBER: US/10/012,231A
; CURRENT FILING DATE: 2002-06-10
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 477
; SEQ ID NO 308
; LENGTH: 671
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-012-231A-308
Query Match 6.3%; Score 143.5; DB 2; Length 671;
Best Local Similarity 24.8%; Pred. No. 0.0013;
Matches 79; Conservative 56; Mismatches 138; Indels 45; Gaps 15;
QY 2 AFQSSPLVKPEASPAESFLS-APGDNFTSLFADSTP-----STLNPRDMMTPD 49
Db 170 ARKASSDLDQASVSPSEENSESSSESEKTSQDQTFPEKAAVAPRRGPLGRKKKKAP 229
QY 50 SVADIDSRL---SVIPESQDAEDDESHSTATAPSTSE---KKPVKKRKSQGVLPPEKPT 103
Db 230 SASDSKADSKADSGAKPEPVAMARSASSSSSSSSSDSDSVSVKKPPRGRKPAEKPLPKPRG 289
QY 104 NLPRKR--AKTEDEKEORRVERVLNRRAAQSSRRERKLEVEALEKRKNELETLLINVQ 161
Db 290 RKPKEPRPSSSSSDSDSDVDRISEWKR-----RDEARRRLEARRRRRQEBELRLREQ 345
QY 162 KTNLILVEELNRRFRSSGVVTRSS--SPLDSL-QDSITLSQQLFGSRDQ-----TWS 211
Db 346 EKE---EKERRRERADRGEAERGSGGSGGDELREDEPEVKR---GRKGRGPPSSSS 399
QY 212 NPEQSLMDQIMRSANPTVNPASLPSLPPISDKFQTKKEDEEQADEDEEQADEDEE 271
Db 400 EPEAELEAEAKKSAKP--QSSSTEPARKP-CQEKRVPRPEKQQA-KPVKVERT-RKRS 454
QY 272 EAAAKEKNSQSRVSTD 289
Db 455 EGFSMDRKVEKKKEPSVE 472
RESULT 9
US-10-015-389A-308
; Sequence 308, Application US/10015389A
; Patent No. 6936436
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan I.
; APPLICANT: Ferrara, Napoleone

```
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth J.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; FILE REFERENCE: P2830P1C48
; CURRENT APPLICATION NUMBER: US/10/015,389A
; CURRENT FILING DATE: 2002-06-25
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 477
; SEQ ID NO 308
; LENGTH: 671
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-015-389A-308

Query Match
Best Local Similarity 24.8%; Pred. No. 0.0013;
Matches 79; Conservative 56; Mismatches 138; Indels 45; Gaps 15;

QY 2 AFQSSPLVKFASPAESFLS-APGDNFTSLPADSTP-----STLNPRDMTDP 49
Db 170 ARKASSDLQASVSPSEENSESSSEKTSQDFTPEKKAAVRAPRRGPLGRKKKAP 229
QY 50 SVADIDSLR---SVIPESQDAEDDHSHTSATPTSE---KKPVKKRSWQVLPKPT 103
Db 230 SASDSKADSGAKPEPVAMARSASSSSSSSDSDSVVKKPPRGRKPAEKPLPKPRG 289
QY 104 NLPPRKR--AKTEDEKEQRRVRLNRRAAQSSRRERKRLVEALEKRNKELETLLINQV 161
Db 290 RKPKPERPPSSSSSDSDSDVDRISEWKR-----RDEARRLEARRRQEBELRLREQ 345
QY 162 KTNLILVELNFRSSGVVTRSS--SPLDSL-QDSITLSQOLFGRDQ-----TMS 211
Db 346 EKE---EKERRERADRGEAERGSGGDELREDDPEVKKR--GRKGRGPPSSSDS 399
QY 212 NPEQSLMQIWRSAANPTVNPASLSPSPPIPSDKFEQTKEDDEEQADEDEMEQTHETK 271
Db 400 EPEALEEREAKSAKKP--QSSSTEPARKP-GQKEKRVPRPEKQQA-KPVKVERT-RKRS 454
QY 272 EAAAKEKNSQSRVSTD 289
Db 455 EGFSMDRKVEKKKPSVE 472

RESULT 10
US-10-005-768A-308
; Sequence 308, Application US/10006768A
; Patent No. 6936597
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan I.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth J.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; FILE REFERENCE: P2830P1C47
; CURRENT APPLICATION NUMBER: US/10/015,671A
; CURRENT FILING DATE: 2001-12-11
; Prior application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 477
; SEQ ID NO 308
; LENGTH: 671
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-015-671A-308

Query Match
Best Local Similarity 24.8%; Pred. No. 0.0013;
Matches 79; Conservative 56; Mismatches 138; Indels 45; Gaps 15;

QY 2 AFQSSPLVKFASPAESFLS-APGDNFTSLPADSTP-----STLNPRDMTDP 49
Db 170 ARKASSDLQASVSPSEENSESSSEKTSQDFTPEKKAAVRAPRRGPLGRKKKAP 229
QY 50 SVADIDSLR---SVIPESQDAEDDHSHTSATPTSE---KKPVKKRSWQVLPKPT 103
Db 230 SASDSKADSGAKPEPVAMARSASSSSSSSDSDSVVKKPPRGRKPAEKPLPKPRG 289
QY 104 NLPPRKR--AKTEDEKEQRRVRLNRRAAQSSRRERKRLVEALEKRNKELETLLINQV 161
Db 290 RKPKPERPPSSSSSDSDSDVDRISEWKR-----RDEARRLEARRRQEBELRLREQ 345
QY 162 KTNLILVELNFRSSGVVTRSS--SPLDSL-QDSITLSQOLFGRDQ-----TMS 211
Db 346 EKE---EKERRERADRGEAERGSGGDELREDDPEVKKR--GRKGRGPPSSSDS 399
QY 212 NPEQSLMQIWRSAANPTVNPASLSPSPPIPSDKFEQTKEDDEEQADEDEMEQTHETK 271
Db 400 EPEALEEREAKSAKKP--QSSSTEPARKP-GQKEKRVPRPEKQQA-KPVKVERT-RKRS 454
QY 272 EAAAKEKNSQSRVSTD 289
Db 455 EGFSMDRKVEKKKPSVE 472

RESULT 11
US-10-015-671A-308
; Sequence 308, Application US/10015671A
; Patent No. 6946263
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan I.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth J.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; FILE REFERENCE: P2830P1C47
; CURRENT APPLICATION NUMBER: US/10/015,671A
; CURRENT FILING DATE: 2001-12-11
; Prior application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 477
; SEQ ID NO 308
; LENGTH: 671
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-015-671A-308

Query Match
Best Local Similarity 24.8%; Pred. No. 0.0013;
Matches 79; Conservative 56; Mismatches 138; Indels 45; Gaps 15;
```

Matches 79; Conservative 56; Mismatches 138; Indels 45; Gaps 15;
QY 2 AFQSSPLVKEASPAESFLS-APGDNFTSLFADSTP-----STLNPRDMWTPD 49
Db 170 ARKASSDLQASVSPSEENSESESEKTSQDQFTPEKKAAVRAPRRGPGJGGRKKKAP 229
QY 50 SVADIDSRL---SVIPESQDAEDDESHSTSATAPSTSE---KKPVKKRKSQGVLPPEPKT 103
Db 230 SASDSKADSDGAKPEPVMARSASSSSSSSSSDSDSVVKPPGRKPAEKPLPKPRG 289
QY 104 NLPPRKR--AKTEDEKEQRRVERVLNRRRAAQSRRERKLEVEALEKRNKELETLLINQV 161
Db 290 RKPKPERPPSSSSSDSDSDVDRISEWKR---RDEARRRELEARRRQOEELRLREQ 345
QY 162 KTNLILVEELNRRSSGGVWTRSS--SPLDSL-QDSITLSQQLFGSRDQ-----TMS 211
Db 346 EKE---EKERRRERADRGEARGSGGSDGDELREDDPEVKR---GRKGRGPPSSSDS 399
QY 212 NPEQSLMDQIMRSAAANTVNPASLSPSPPIISDKFQTKKEDEQADEDEMEQTHWHTK 271
Db 400 EPEAELEREAKSAKKP--QSSSTEPARKP-GQKEKRVPRPEKQQA-KPVKVERT-RKRS 454
QY 272 EAAAAKEKNKQSRVSTD 289
Db 455 EGFSMDRKVKKKEPSVE 472
RESULT 12
US-10-015-393A-308
; Sequence 308, Application US/10015393A
; Patent No. 6951737
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Botstein, David
; APPLICANT: Eaton, Dan I.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Fong, Sherman
; APPLICANT: Goddard, Audrey
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth J.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; FILE REFERENCE: P2830P1C46
; CURRENT FILING DATE: 2002-06-10
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 477
; SEQ ID NO 308
; LENGTH: 671
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-015-393A-308
Query Match 6.3%; Score 143.5; DB 2; Length 671;
Best Local Similarity 24.8%; Pred. No. 0.0013;
Matches 79; Conservative 56; Mismatches 138; Indels 45; Gaps 15;
QY 2 AFQSSPLVKEASPAESFLS-APGDNFTSLFADSTP-----STLNPRDMWTPD 49
Db 170 ARKASSDLQASVSPSEENSESESEKTSQDQFTPEKKAAVRAPRRGPGJGGRKKKAP 229
QY 50 SVADIDSRL---SVIPESQDAEDDESHSTSATAPSTSE---KKPVKKRKSQGVLPPEPKT 103
Db 230 SASDSKADSDGAKPEPVMARSASSSSSSSSSDSDSVVKPPGRKPAEKPLPKPRG 289
QY 104 NLPPRKR--AKTEDEKEQRRVERVLNRRRAAQSRRERKLEVEALEKRNKELETLLINQV 161
Db 290 RKPKPERPPSSSSSDSDSDVDRISEWKR---RDEARRRELEARRRQOEELRLREQ 345
QY 162 KTNLILVEELNRRSSGGVWTRSS--SPLDSL-QDSITLSQQLFGSRDQ-----TMS 211
Db 346 EKE---EKERRRERADRGEARGSGGSDGDELREDDPEVKR---GRKGRGPPSSSDS 399
QY 212 NPEQSLMDQIMRSAAANTVNPASLSPSPPIISDKFQTKKEDEQADEDEMEQTHWHTK 271
Db 400 EPEAELEREAKSAKKP--QSSSTEPARKP-GQKEKRVPRPEKQQA-KPVKVERT-RKRS 454
QY 272 EAAAAKEKNKQSRVSTD 289
Db 455 EGFSMDRKVKKKEPSVE 472

Db 290 RKPKPERPPSSSSSDSDSDVDRISEWKR---RDEARRRELEARRRQOEELRLREQ 345
QY 162 KTNLILVEELNRRSSGGVWTRSS--SPLDSL-QDSITLSQQLFGSRDQ-----TMS 211
Db 346 EKE---EKERRRERADRGEARGSGGSDGDELREDDPEVKR---GRKGRGPPSSSDS 399
QY 212 NPEQSLMDQIMRSAAANTVNPASLSPSPPIISDKFQTKKEDEQADEDEMEQTHWHTK 271
Db 400 EPEAELEREAKSAKKP--QSSSTEPARKP-GQKEKRVPRPEKQQA-KPVKVERT-RKRS 454
QY 272 EAAAAKEKNKQSRVSTD 289
Db 455 EGFSMDRKVKKKEPSVE 472
RESULT 13
US-10-011-833A-308
; Sequence 308, Application US/10011833A
; Patent No. 6951920
; GENERAL INFORMATION:
; APPLICANT: Baker, Kevin P.
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan I.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth J.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; FILE REFERENCE: P2830P1C22
; CURRENT FILING DATE: 2002-06-25
; Prior Application removed - See File Wrapper or Palm
; NUMBER OF SEQ ID NOS: 477
; SEQ ID NO 308
; LENGTH: 671
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-011-833A-308
Query Match 6.3%; Score 143.5; DB 2; Length 671;
Best Local Similarity 24.8%; Pred. No. 0.0013;
Matches 79; Conservative 56; Mismatches 138; Indels 45; Gaps 15;
QY 2 AFQSSPLVKEASPAESFLS-APGDNFTSLFADSTP-----STLNPRDMWTPD 49
Db 170 ARKASSDLQASVSPSEENSESESEKTSQDQFTPEKKAAVRAPRRGPGJGGRKKKAP 229
QY 50 SVADIDSRL---SVIPESQDAEDDESHSTSATAPSTSE---KKPVKKRKSQGVLPPEPKT 103
Db 230 SASDSKADSDGAKPEPVMARSASSSSSSSSSDSDSVVKPPGRKPAEKPLPKPRG 289
QY 104 NLPPRKR--AKTEDEKEQRRVERVLNRRRAAQSRRERKLEVEALEKRNKELETLLINQV 161
Db 290 RKPKPERPPSSSSSDSDSDVDRISEWKR---RDEARRRELEARRRQOEELRLREQ 345
QY 162 KTNLILVEELNRRSSGGVWTRSS--SPLDSL-QDSITLSQQLFGSRDQ-----TMS 211
Db 346 EKE---EKERRRERADRGEARGSGGSDGDELREDDPEVKR---GRKGRGPPSSSDS 399
QY 212 NPEQSLMDQIMRSAAANTVNPASLSPSPPIISDKFQTKKEDEQADEDEMEQTHWHTK 271
Db 400 EPEAELEREAKSAKKP--QSSSTEPARKP-GQKEKRVPRPEKQQA-KPVKVERT-RKRS 454
QY 272 EAAAAKEKNKQSRVSTD 289
Db 455 EGFSMDRKVKKKEPSVE 472

Db 455 EGFSMDRKVEKKEPSVE 472

RESULT 14

US-10-006-041A-308

; Sequence 308, Application US/10006041A

; Patent No. 6951921

; GENERAL INFORMATION:

; APPLICANT: Baker, Kevin P.

; APPLICANT: Botstein, David

; APPLICANT: Desnoyers, Luc

; APPLICANT: Eaton, Dan I.

; APPLICANT: Ferrara, Napoleone

; APPLICANT: Fong, Sherman

; APPLICANT: Gao, Wei-Qiang

; APPLICANT: Goddard, Audrey

; APPLICANT: Godowski, Paul J.

; APPLICANT: Grimaldi, Christopher J.

; APPLICANT: Gurney, Austin L.

; APPLICANT: Hillan, Kenneth J.

; APPLICANT: Pan, James

; APPLICANT: Paoni, Nicholas F.

; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic

; FILE REFERENCE: P2830PIC8

; CURRENT APPLICATION NUMBER: US/10/006,041A

; Prior Filing Date: 2001-12-06

; Prior Application removed - See File Wrapper or Palm

; NUMBER OF SEQ ID NOS: 477

; SEQ ID NO 308

; LENGTH: 671

; TYPE: PRT

; ORGANISM: Homo sapiens

US-10-006-041A-308

Query Match 6.3%; Score 143.5; DB 2; Length 671;

Best Local Similarity 24.8%; Pred. No. 0.0013;

Matches 79; Conservative 56; Mismatches 138; Indels 45; Gaps 15;

QY 2 AFQSSPLVKFEASPASFLS-APGDNFTSLPADSTP-----STLNPRDMMTDP 49

Db 170 ARKASSDLQASVSPSEENSESEKTSQDFTPEKKAAPRRGPGKPKKKKAP 229

QY 50 SVADIDSLRL---SVIPESQDAEDDESHSTSATAPSTSE---KKPVKKRKSQGVLPPEPT 103

Db 230 SASDSKADSDGAKPEPVMARSASSSSSSSSSDSDSVSKPPRGKPAEKPLPKPRG 289

QY 104 NLPPKR--AKTEDEKQRRVRLNRRAAQSSRRKRLEVEALEKKNKELETLLINVQ 161

Db 290 RKPKPERPPSSSSSDSDSDVDRISEWKR-----RDEARRRLEARRRREQEELRLREQ 345

QY 162 KTNLLILVEELNRRSSGVVTRSS--SPLDSL-ODSITLSQQLFGSRDQ-----TMS 211

Db 346 EKE---EKERRRERADRGEAERGSGGSDGDELREDEPVKKR--GRKGRGPGPPSSSDS 399

QY 212 NPEQSLMDQIMRSAANTVNPASLSPSLPTSDKFEQTKEDDEBQADDEMEQTHETK 271

Db 400 EPEALEEREAKKSARKP--QSSSTEPARKP-GQKEKRVPEEKQQA-KPVKVERT-RKRS 454

QY 272 EAAAKEKNKSKQSRVSTD 289

Db 455 EGFSMDRKVEKKEPSVE 472

RESULT 15

US-10-012-064A-308

; Sequence 308, Application US/10012064A

; Patent No. 6953841

; GENERAL INFORMATION:

; APPLICANT: Baker, Kevin P.

; APPLICANT: Botstein, David

; APPLICANT: Desnoyers, Luc

; APPLICANT: Eaton, Dan I.

Search completed: November 23, 2005, 03:27:32

; APPLICANT: Ferrara, Napoleone
; APPLICANT: Fong, Sherman
; APPLICANT: Gao, Wei-Qiang
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, Christopher J.
; APPLICANT: Gurney, Austin L.
; APPLICANT: Hillan, Kenneth J.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; FILE REFERENCE: P2830PIC19
; CURRENT APPLICATION NUMBER: US/10/012,064A
; CURRENT FILING DATE: 2002-07-15
; PRIOR APPLICATION NUMBER: 60/098716
; PRIOR FILING DATE: 1998-09-01
; PRIOR APPLICATION NUMBER: 60/098723
; PRIOR FILING DATE: 1998-09-01
; PRIOR APPLICATION NUMBER: 60/098749
; PRIOR FILING DATE: 1998-09-01
; PRIOR APPLICATION NUMBER: 60/098750
; PRIOR FILING DATE: 1998-09-01
; PRIOR APPLICATION NUMBER: 60/098803
; PRIOR FILING DATE: 1998-09-02
; PRIOR APPLICATION NUMBER: 60/098821
; PRIOR FILING DATE: 1998-09-02
; PRIOR APPLICATION NUMBER: 60/098843
; PRIOR FILING DATE: 1998-09-02
; PRIOR APPLICATION NUMBER: 60/099536
; PRIOR FILING DATE: 1998-09-09
; PRIOR APPLICATION NUMBER: 60/099596
; PRIOR FILING DATE: 1998-09-09
; PRIOR APPLICATION NUMBER: 60/099598
; PRIOR FILING DATE: 1998-09-09
; Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 477
; SEQ ID NO 308
; LENGTH: 671
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-012-064A-308

Query Match 6.3%; Score 143.5; DB 2; Length 671;

Best Local Similarity 24.8%; Pred. No. 0.0013;

Matches 79; Conservative 56; Mismatches 138; Indels 45; Gaps 15;

QY 2 AFQSSPLVKFEASPASFLS-APGDNFTSLPADSTP-----STLNPRDMMTDP 49

Db 170 ARKASSDLQASVSPSEENSESEKTSQDFTPEKKAAPRRGPGKPKKKKAP 229

QY 50 SVADIDSLRL---SVIPESQDAEDDESHSTSATAPSTSE---KKPVKKRKSQGVLPPEPT 103

Db 230 SASDSKADSDGAKPEPVMARSASSSSSSSSSDSDSVSKPPRGKPAEKPLPKPRG 289

QY 104 NLPPKR--AKTEDEKQRRVRLNRRAAQSSRRKRLEVEALEKKNKELETLLINVQ 161

Db 290 RKPKPERPPSSSSSDSDSDVDRISEWKR-----RDEARRRLEARRRREQEELRLREQ 345

QY 162 KTNLLILVEELNRRSSGVVTRSS--SPLDSL-ODSITLSQQLFGSRDQ-----TMS 211

Db 346 EKE---EKERRRERADRGEAERGSGGSDGDELREDEPVKKR--GRKGRGPGPPSSSDS 399

QY 212 NPEQSLMDQIMRSAANTVNPASLSPSLPTSDKFEQTKEDDEBQADDEMEQTHETK 271

Db 400 EPEALEEREAKKSARKP--QSSSTEPARKP-GQKEKRVPEEKQQA-KPVKVERT-RKRS 454

QY 272 EAAAKEKNKSKQSRVSTD 289

Db 455 EGFSMDRKVEKKEPSVE 472

Job time : 46.1647 secs



Protein Sequence Searches - February 2005

All of the sequence databases on ABSS have recently been updated.

- Please note that the curators of the UniProt database have purged some temporary accession numbers from the most recent version of UniProt. These sequences have been assigned new permanent accession numbers. The new UniProt record may not contain the previous temporary accession number.
- If you encounter an accession number from an older search run against UniProt (results file extension **.rup**) that can no longer be found in the database, the permanent record with the new accession number can be found by searching the old accession number in the UniProt Protein Archive database (UniPARC) at:

<http://www.pir.uniprot.org/database/archive.shtml>

If you have any questions regarding this information or your results, please contact any STIC searcher.

When submitting sequence search results for scanning into IFW, please include a copy of this attachment to assist any future Examiners or members of the public who may encounter UniProt temporary accession numbers.

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 23, 2005, 03:05:10 ; Search time 193.805 Seconds
(without alignments)
1641.819 Million cell updates/sec

Title: US-10-663-450-2
Perfect score: 2294
Sequence: 1 MAFOQSSPLVKFEAPAESF.....QFQSGASHGCGDDGGIAGV 451

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Uniprot 05.80.*
1: uniprot_sprot.*
2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB	ID	Description
1	2294	100.0	451	2	Q8TF33_TRIRE	Q8tf33 trichoderma
2	885	38.6	429	2	Q4HTT5_GIBZE	Q4ht5 gibberella
3	756	33.0	430	2	Q7SHF0_NEUCR	Q7shf0 neurospora
4	587	25.6	556	2	Q51KW8_MAGGR	Q51kw8 magnaporthe
5	437.5	19.1	350	2	Q8TFU8_EMENI	Q8tfu8 emericeella
6	426.5	18.6	342	2	Q6W8X2_ASPEG	Q6w8x2 aspergillus
7	424	18.5	347	2	Q5AQN3_EMENI	Q5aqn3 aspergillus
8	342	14.9	433	2	Q4WY88_ASPPU	Q4wey8 aspergillus
9	285	12.4	357	2	Q5AA52_CANAL	Q5aa52 candida alb
10	230.5	10.0	299	2	Q6CBV1_YARLI	Q6cbv1 yarrowia li
11	211	9.2	230	1	HAC1_YEAST	P41546 saccharomyc
12	208.5	9.1	260	2	Q6BQC2_DEBHA	Q6bc2 debaryomyc
13	206.5	9.0	273	2	Q6CKQ1_KLUUA	Q6ckq1 kluyveromyc
14	197	8.6	329	2	Q6FLY3_CANGA	Q6fly3 candida gla
15	190	8.3	228	2	Q75BQ5_ASHGO	Q75bq5 ashbya goes
16	172.5	7.5	1225	2	Q4P0Y9_USTMA	Q4p0y9 ustilago ma
17	172	7.5	595	2	Q4P8Q4_USTMA	Q4p8q4 ustilago ma
18	170.5	7.4	690	2	Q44743_CAEBL	Q44743 caenorhabdi
19	169.5	7.4	588	2	Q411V8_GIBZE	Q411v8 gibberella
20	166	7.2	3257	2	Q9V736_DROME	Q9v736 drosophila
21	164.5	7.2	1444	2	Q9VTN2_DROME	Q9vtn2 drosophila
22	164.5	7.2	1514	2	Q8SY55_DROME	Q8sy55 drosophila
23	163.5	7.1	1196	2	Q4SUB7_TETNG	Q4sub7 tetraodon n
24	162	7.1	615	2	Q4W8H0_ASPPU	Q4w8h0 aspergillus
25	159.5	7.0	2938	2	Q4IR68_GIBZE	Q4ir68 gibberella
26	157.5	6.9	868	2	Q4P172_USTMA	Q4p172 ustilago ma
27	157	6.8	1214	2	Q4W3T0_ASPPU	Q4w3t0 aspergillus
28	156.5	6.8	782	2	Q870C5_9SACH	Q870c5 kluyveromyc
29	156.5	6.8	3340	2	Q4QB22_LEIMA	Q4qbb2 leishmania
30	155.5	6.8	2041	2	Q7S374_NEUCR	Q7s374 neurospora
31	155	6.8	583	1	API_KLUUA	P56095 kluyveromyc

32	155	6.8	7480	2	Q4IP09_GIBZE	Q4ip09 gibberella
33	154.5	6.7	737	2	Q5KPM1_CRYNE	Q5kpm1 cryptococcu
34	154	6.7	947	2	Q20625_CABEL	Q20625 caenorhabdi
35	153.5	6.7	1357	2	Q4P8Z5_USTMA	Q4p8z5 ustilago ma
36	153.5	6.7	2062	2	Q9JYJ5_MOUSE	Q9jyj5 mus musculu
37	153	6.7	707	2	Q569T3_XENLA	Q569t3 xenopus lae
38	153	6.7	704	2	Q6C373_YARLI	Q6c373 yarrowia li
39	153	6.7	912	2	Q9XSC3_BOVIN	Q9xsc3 bos taurus
40	153	6.7	1230	2	Q54JY4_DICDI	Q54jy4 dicystosteli
41	152.5	6.6	686	2	Q61D33_CAEBR	Q61d33 caenorhabdi
42	152.5	6.6	1450	2	Q80TR9_MOUSE	Q80tr9 mus musculu
43	152	6.6	709	2	Q560M5_CRYNE	Q560m5 cryptococcu
44	151.5	6.6	788	2	Q6DE25_XENLA	Q6de25 xenopus lae
45	151.5	6.6	937	2	Q6C621_YARLI	Q6c621 yarrowia li

ALIGNMENTS

RESULT 1
Q8TF33 TRIRE PRELIMINARY; PRT; 451 AA.
AC Q8TF33;
DT 01-JUN-2002 (TREMBlrel. 21, Created)
DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)
DE Transcription factor.
GN Name=hacl;
OS Trichoderma reesei (Hypocrea jecorina).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Hypocreomycetidae; Hypocreales; Hypocreaceae; Hypocrea.
OX NCBI_Taxid=51453;
RN [1]
NUCLEOTIDE SEQUENCE.
RA Salohelmo M.L.A., Valkonen M., Penttilae M.E.;
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; A413272; CAC8374.1; -; Genomic_DNA.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR011700; bZIP 2.
DR InterPro; IPR004827; TF_bZIP.
DR Pfam; PF07716; bZIP_2; 1.
DR SMART; SM00338; BRLZ; 1.
DR PROSITE; PS50217; BZIP; 1.
KW DNA-binding; Nuclear protein.
SQ SEQUENCE 451 AA; 49277 MW; 600F10B471EA3AD3 CRC64;

Query Match	100.0%;	Score	2294;	DB	2;	Length	451;
Best Local Similarity	100.0%;	Pred. No.	2.8e-104;	Mismatches	0;	Indels	0;
Matches	451;	Conservative	0;	0;	0;	Gaps	0;
QY	1	MAFOQSSPLVKFEAPAESF	LSAPGDNFTSLFADSTPSTLNPRDMMTPSDVADISRLSV	60			
Db	1	MAFOQSSPLVKFEAPAESF	LSAPGDNFTSLFADSTPSTLNPRDMMTPSDVADISRLSV	60			
QY	61	IPESQDAEDDESHSATSAPSTSEKKPVKKRSGQVLPEPKTNLPKRKAKTDEKEQR	120				
Db	61	IPESQDAEDDESHSATSAPSTSEKKPVKKRSGQVLPEPKTNLPKRKAKTDEKEQR	120				
QY	121	RVERVLNRAAQSSRRKRLVEALEKENKLETLLINVQKTNLLVVELNRFRESSGV	180				
Db	121	RVERVLNRAAQSSRRKRLVEALEKENKLETLLINVQKTNLLVVELNRFRESSGV	180				
QY	181	VTRSSPLDLSQDSTTLSQOLFSGRDGQTMSPNEQSLMDQIMRSAANPTVNPASLSPSLP	240				
Db	181	VTRSSPLDLSQDSTTLSQOLFSGRDGQTMSPNEQSLMDQIMRSAANPTVNPASLSPSLP	240				
QY	241	PISDKFEQTKEDDEEQADEDEMEQTHETKEAAAAKKNQSKQSVSTSTQRPVAVSIGG	300				
Db	241	PISDKFEQTKEDDEEQADEDEMEQTHETKEAAAAKKNQSKQSVSTSTQRPVAVSIGG	300				
QY	301	DAVPFSDDAAGANCLGLDPVHQDDGPPFSGHISFGLSAAALDADRYLLESQLLASPNASTV	360				

```

116 DESTDTLNRKRAKTEDEKEQRRVERVLNRRAAQSRRERKQVEALEKRNQSELEAFMA 175
160 VOKTNLILVEELNRRFRSSGVVTRSSPLDSQDSITLSQQLFGSRDGTMSNPEQSLMD 219
176 AQENAKLMSELEQIR-SCAVSYSPVLDSPRVNPSLSQELFGSKH---THEPIDGLV- 230
220 QIMRSAANPTVNPASLSPSPISDKFQTKBEDEBEQADEDEMEQTHWETKEAAAAKEK 279
231 -----SSNTVDPTALSPVLSVAESFEBIAEQEPSNEAKPBLTEST----- 272
280 NSKQSRVSTDTORPAVSGDAAVPVFSDDAGANCLGLDPVHQDDGPPSIGHSGFLSAA 339
273 -----SPDLTQ--LSQVGDAAQVPSAANLDAANLGLAPALPGDAAFSLGNSDILPTS 323
340 LDADRYLLESQLLASPNASTVDDVYLAGDSAACTNPLPSDYPDFDINDFLTDANHAAYD 399
324 IGADRYIILENKYLSDDSSIIGDDVMVGDAFANLND-----DFDLSLWLNDDSAISA-E 377
400 IVAASNYAAADRELDLEIHPENQIPSRHSIQOQSGASSHCGCDGCGIAGV 451
378 SMATSDFAAAIQGLEPKIYEPENQVSSENPIQOPHPGASTQCGDVGGIAGV 429

RESULT 3
Q7SHFO_NEUCR PRELIMINARY; PRT; 430 AA.
ID Q7SHFO1;
AC Q7SHFO1;
DT 01-MAR-2004 (TREMELrel. 26, Created)
DT 01-MAR-2004 (TREMELrel. 26, Last sequence update)
DE 01-MAR-2004 (TREMELrel. 26, Last annotation update)
DE Predicted protein.
GN Name=NCU01856.1;
OS Neurospora crassa.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.
OC NCBI_TaxID=5141;
RN [1]
RC NUCLEOTIDE SEQUENCE.
RC STRAIN=OR74A;
RA Galagan J.E., Calvo S.E., Borkovich K.A., Selker E.U., Read N.D.,
RA Jaffé D., FitzHugh W., Ma L.-J., Smirnov S., Purcell S., Rehman B.,
RA Elkins T., Engels R., Wang S., Nielsen C.B., Butler J., Endrizzi M.,
RA Qui D., Ianakiev P., Pedersen D., Nelson M., Washburne M.,
RA Selitrnmlkoff C.P., Kinsey J.A., Braun E.L., Zelter A., Schulte U.,
RA Kothe G.O., Jedd G., Mewes W., Staben C., Marcotte E., Greenberg D.,
RA Roy A., Foley K., Naylor J., Thomann N., Barrett R., Gnerre S.,
RA Kamal M., Kamyselis M., Mauceli E., Bielke C., Rudd S., Frishman D.,
RA Krystofova S., Rasmussen C., Metzberg R.L., Perkins D.D., Kroken S.A.,
RA Cogoni C., Macino G., Catchside D., Li W., Pratt R.J., Osmari S.A.,
RA DeSouza C.C., Glass L., Orbach M.J., Berglund J., Voelker R.,
RA Yarden O., Plamann M., Seiler S., Dunlap J., Radford A., Aramayo R.,
RA Natvig D.O., Alex L.A., Mannhaupt G., Ebbola D.J., Freitag M.,
RA Paulsen I., Sachs M.S., Lander E.S., Nuebaum C., Birren B.;
RT "The Genome Sequence of the Filamentous Fungus Neurospora crassa.";
RL Nature 0:0-0(2003).
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; AABX01000004; EAA36251.1; -; Genomic_DNA.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR011700; bZIP 2.
DR Pfam; PF0716; bZIP_2; 1.
DR PROSITE; PS50217; bZIP; 1.
SQ SEQUENCE 430 AA; 46599 MW; 0EE0657CFA6160DA CRC64;

Query Match 33.0%; Score 756; DB 2; Length 430;
Best Local Similarity 42.3%; Pred. No. 2, 8e-29;
Matches 197; Conservative 52; Mismatches 157; Indels 60; Gaps 14;

QY 2 AFQSSPLVKFEAPSEFLSAPGDNFTSLFADSTPSTLNPR-DMWTPDSPADIDSRLSV 60

```

```

Db 6 AHSPANLKE-NPASLLATPGIPEVFGSDRATSPSLDVMSPVSLAD-DVDLTA 63
Qy 61 I-----PESQDAEDDESHSTSATAPSTSEKKPVKRGKSGQVLPPEKTNLPPPKRA 111
Db 64 LATLVQIPRSTPA-----STPAPET-EKKPVKRGKSGQVLPPEKTNLPPPKRA 114
Qy 112 KTEDEKQRRVRLNRRAAQSRRERKRLVEALEKRNKELETLLINVOKTNLILVEEL 171
Db 115 KTEDEKQRRVRLNRRAAQSRRERKRLVEALEKRNKELETLLINVOKTNLILVEEL 174
Qy 172 NRPRSSGVTRSSPLDSLQDS-ITLSQQLFGSRDGTMSNPQSLMDQIMRSAANPTV 230
Db 175 RENGVAITATRPAS-FDGLNPTVTTSQELFSSQGHNLKHDSSLEQLFPYIKTEETV 233
Qy 231 NPASLSPPLPISDKFQTKKEDEBEQADEMEQWTWHEKRAAAKRNKSKRSRVSTDS 290
Db 234 NPASLSPVNLPLPME-----EDGEKEQS---TAAQPVADATSTATVDTSPDA 278
Qy 291 TORPAVSIIGDAAVPVFDGANGCLGLDPVHQDGFSGHSGFLSAAALDADRYLLESQ 350
Db 279 TOHPAVVFPVAVSAPVAGSSLSAAPLGLGNFMDD-----LVLSKTFDADRYIYETE 329
Qy 351 LLASPNASTVDDDLVLAGDSAACTNPLPSVDYDPDINDFLTDDANHAAYDIVAASNYAAD 410
Db 330 FFSPPSPNSNDEVMAGDGTETLHPANPENFPD--BFINPDVSN-----ANSNEQQPQ 382
Qy 411 RE-----LDLEIHDPENIPSRHSIQPOSGASSHCGDDGGIAGV 450
Db 383 QQSQFCATTNASVHNLTPNASEDPYQOPHTGASLNGCDGGIAGV 428

RESULT 4
Q51KW8 MAGGR PRELIMINARY; PRT; 556 AA.
ID Q51KW8 MAGGR PRELIMINARY; PRT; 556 AA.
AC Q51KW8;
DT 13-SEP-2005 (Tremblrel. 31, Created)
DT 13-SEP-2005 (Tremblrel. 31, Last sequence update)
DE Hypothetical protein.
GN ORFNames=MG09010.4;
OS Magnaporthe grisea 70-15.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariomycetes incertae sedis; Magnaporthaceae; Magnaporthe.
OX NCBI_TaxID=242507;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=70-15;
RA Birren B., Nusbau C., Abebe A., Abouelleil A., Adekoya E.,
RA Alt-zahra M., Allen N., Allen T., An P., Anderson M., Anderson S.,
RA Arachchi H., Armbruster J., Bachanteang P., Baldwin J., Barry A.,
RA Bayul T., Blitshsteyn B., Bloom T., Blye J., Boguslavskiy L.,
RA Borowsky M., Boukhgalter B., Brunache A., Butler J., Calixte N.,
RA Calvo S., Camarata J., Campo K., Chang J., Cheshatsang Y., Citroen M.,
RA Callimore A., Considine T., Cook A., Cooke P., Corum B., Cuomo C.,
RA David R., Dawoe T., Degray S., Dodge S., Dooley K., Dorje P.,
RA Dorjee K., Dorris L., Duffey N., Dupes A., Elkins T., Engels R.,
RA Erickson J., Farina A., Faro S., Ferreira P., Fischer H.,
RA Fitzgerald M., Foley K., Gage D., Galagan J., Gearin G., Gnerre S.,
RA Girke A., Goyette A., Graham J., Grandbois E., Gyaltsen K., Hafez N.,
RA Honan T., Horn A., Houde N., Hughes L., Hulme W., Husby E., Iliev I.,
RA Jaffe D., Jones C., Kamal M., Kamat A., Kamvesselis M., Karlsson E.,
RA Kells C., Kieu A., Kiser P., Kodira C., Kulbokas E., Labutti K.,
RA Lama D., Landers T., Leger J., Levine S., Lewis D., Lewis T.,
RA Lindblad-toh K., Liu X., Lokyitsang T., Lokyitsang Y., Lucien O.,
RA Lui A., Ma L.J., Mabbitt R., Macdonald J., Maclean C., Major J.,
RA Manning J., Marabella R., Maru K., Matthews C., Maucelli E.,
RA McCarthy M., McDonough S., McGhee T., Meldrum J., Meneus L.,
RA Mesirov J., Minhalin A., Mihova T., Mikkelson T., Mlenga V., Moru K.,
RA Mozes J., Mulrain L., Munson G., Naylor J., Neues C., Nguyen C.,
RA Nguyen N., Nguyen T., Nicol R., Nielsen C., Nizzari M., Norbu C.,
RA Norbu N., O'donnell P., Okoawo O., O'leary S., Omotosho B.,

```

```

RA O'Neill K., Osman S., Parker S., Perrin D., Phunkhang P., Piquani B.,
RA Purcell S., Rachupka T., Ramasamy U., Rameau R., Ray V., Raymond C.,
RA Retta R., Richardson S., Rise C., Rodriguez J., Rogers J., Rogov P.,
RA Rutman M., Schubach R., Seaman C., Settupalli S., Sharpe T.,
RA Sheridan J., Sherpa N., Shi J., Smirnov S., Smith C., Sougnez C.,
RA Spencer B., Stalker J., Stange-thomann N., Stavropoulos S.,
RA Stetson K., Stone C., Stone S., Stubbs M., Talamas J., Tchuinga P.,
RA Tenzing P., Tesfaye S., Theodore J., Thoulutseang Y., Topham K.,
RA Towey S., Teamlia T., Tsomo N., Vallie D., Vassiliev H.,
RA Venkataraman V., Vinson J., Vo A., Wade C., Wang S., Wangchuk T.,
RA Wangdi T., Whittaker C., Wilkinson J., Wu Y., Wyman D., Yadav S.,
RA Yang S., Yang X., Yeager S., Yee E., Young G., Zainoun J., Zembeck L.,
RA Zimmer A., Zody M., Zander E.;
RT "The genome sequence of Magnaporthe grisea.";
RL Submitted (OCT-2003) to the EMBL/GenBank/DBDJ databases.
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=70-15;
RA Dean R., Mitchell T., Brown D., Pan H., Thon M.;
RL Submitted (OCT-2003) to the EMBL/GenBank/DBDJ databases.
RN [3]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=70-15;
RA Zhu H., Blackmon B.;
RL Submitted (OCT-2003) to the EMBL/GenBank/DBDJ databases.
CC -I- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBDJ whole genome shotgun (WGS) entry which is
CC preliminary data.
CC EMBL; AACU01001606; BAA47880.1; -; Genomic_DNA.
DR InterPro; IPR011700; bZIP 2.
DR InterPro; IPR004827; TF_bZIP.
DR Pfam; PF07716; bZIP2; 1.
DR SMART; SM00338; BRLZ; 1.
DR PROSITE; PS0217; bZIP; 1.
KW DNA-binding; Hypothetical protein; Nuclear protein.
SQ SEQUENCE 556 AA; 60798 MW; CFDC4C8D2DC4963C CRC64;

Query Match 25.6%; Score 587; DB 2; Length 556;
Best Local Similarity 45.0%; Pred. No. 6.6e-21;
Matches 148; Conservative 34; Mismatches 93; Indels 54; Gaps 9;

Qy 6 SSPLVKFEASPAESPLSAPGDNFTSLFADTSTLTNPRDMWTPDSVADTDSLSVT-PES 64
Db 6 STPELMEASPAESPLSTEGDMYPSLFGDAKSFSLNPSDLASPGSQEDPVELDAKTES 65
Qy 65 QDAEDDESHST--SATAPSTSEKKPVKRGKSGQVLPPEKTNLPPPKRAKTEDEKQRRV 122
Db 66 AALRAVESTPTPEETSSPAAASEKKTKKRSNGQVLPPEKTNLPPPKRAKTEDEKQRRV 125
Qy 123 ERLVLRRAAQSRRERKRLVEALEKRNKELETLLINVOKTNLILVEELNRPRSSGVVT 182
Db 126 ERLVLRRAAQSRRERKRLVEALEKRNKELETLLINVOKTNLILVEELNRPRSSGVVT 185
Qy 183 RSSPLDSLQDSITLSQQLFGSRD-----GOTMSNPQSLMDQIMRSAANP--TVNP 232
Db 186 SSEKP-----DGLTLPFLRSQDAIATSAATSVSATLADPANSLED-IPRSTNPTVNP 239
Qy 233 ASLSPLPISDKFQTKKEDEBEQADEMEQWTWHEKRAAAKRNKSKRSRVSTDSQ 292
Db 240 ASIRSPAP-----EQHESVASGEE-----AKTSADLTQ 269
Qy 293 RPAVSIIGD-----AAVPVFDGANGCL 316
Db 270 HPAEMLCITDLCQSAKVPQASSQMPASMM 298

RESULT 5
Q8TFUB EMENI PRELIMINARY; PRT; 350 AA.
ID Q8TFUB EMENI PRELIMINARY; PRT; 350 AA.
AC Q8TFUB;
DT 01-JUN-2002 (Tremblrel. 21, Created)
DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)
DT 01-MAR-2004 (Tremblrel. 26, Last annotation update)

```

[illegible]


```

10-MAY-2005 (TrEMBLrel. 30, Created)
10-MAY-2005 (TrEMBLrel. 30, Last sequence update)
10-MAY-2005 (TrEMBLrel. 30, Last annotation update)
DE Hypothetical protein HAC1.
CN Name=HAC1; ORNames=CaO19.2432, CaO19.9968;
OS Candida albicans SC5314.
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX NCBI_TaxID=237561;
RN [1]_
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=SC5314;
RX PubMed=15123810; DOI=10.1073/pnas.0401648101;
RA Jones T., Federspiel N.A., Chibana H., Dungan J., Kalman S.,
RA Magee B.B., Newport G., Thorstenson Y.R., Agabian N., Magee P.T.,
RA Davis R.W., Scherer S.;
RT "The diploid genome sequence of Candida albicans.";
RL Proc. Natl. Acad. Sci. U.S.A. 101:7329-7334 (2004).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=SC5314;
RA Dungan J., Kuo A., Newport G., Lan C.-Y., Iijima C., Adegbola O.,
RA Roberts J., Person K., Donnelly S., Favoreto S., Tzung K.-W.,
RA Jones T., Scherer S., Agabian N.;
RT "Annotation of the Genome of Candida albicans.";
RL Submitted (APR-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; AACQ01000041; EAK99529.1; -; Genomic DNA.
DR EMBL; AACQ01000040; EAK99617.1; -; Genomic DNA.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
KW DNA-binding; Hypothetical protein; Nuclear protein.
SQ SEQUENCE 357 AA; 40127 NW; 8E1C633A2B4ECDAAE CRC64;

Query Match 12.4%; Score 285; DB 2; Length 357;
Best Local Similarity 24.4%; Pred. No. 2.2e-06;
Matches 105; Conservative 67; Mismatches 127; Indels 132; Gaps 15;

QY 31 LEADSTPTLNPRDM--MTPDSVADIDSRSLVSPESQDAEDDESHSTGATPSTSEKPV 88
DB 3 LTVDMNTTSNDDLSVATPTSLM-----TSITTSFMS----- 36
QY 89 KKRKSGWGVL-EP---KTNLPKRKATDEKEQRRVERVLNRRAAQSSRRKRLEVE 144
DB 37 TSTSSHSNTLDDPATFKSTLPPKRKATQEEKEQKIERILNRRAAHASREKKRKEVE 96
QY 145 ALEKRNKELETLINVKTNLILVEELNPRSSGVVTRSS-----SPLDSIQDSITLSQ 199
DB 97 YLENYVLKLET---NLMKLN-----NNYNQAFELLTKNQELLKLEVLDDVSDLKE 146
QY 200 QLFGSRDGTMGNPEQSLMDQIMRGAANPTVNPASLPSLPISDKFEQTKDEEEDQADE 259
DB 147 QTHSNWSGTRSHNKS-----NDEDEEEDDEHQEG 180
QY 260 DEEMEQTWHE-----TKEAATAKESKQSRVSTDSRPAVSTGGDAAPVPSD 309
DB 181 HVEKQEIKEEPVSKRKLNTKSKTKTKSNKMTSTPPSVSLSL-----PDVTN 233
QY 310 DAGANCLGLDPVHQDDGPPSGHSGLSAALADRYLLESQSLASPNASTVDD----- 363
DB 234 FSGTNTWSSPIQ-----IKKEFNIDNIFIKKEFSQSPLOTHOPEQOEOH 280
QY 364 --YLADGSAACFTN-----PLPSDYDFDINDFLTDDA-----NHAAYDI 400
DB 281 EYLYKESKDAFWNPSPSLFSDSPLQIDITSSSSSTSPSSSSLSISSGGTNNHSIADL 340
QY 401 VAASNYAAADR 411
DB 341 AAISLASRANR 351

RESULT 10
Q6CEV1_YARLI
```

```

ID AC Q6CEV1_YARLI PRELIMINARY; PRT; 299 AA.
DT 25-OCT-2004 (TrEMBLrel. 28, Created)
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE Yarrowia lipolytica chromosome B of strain CLIB99 of Yarrowia
DE lipolytica.
GN OrderedLocusNames=VALI0B12716g;
OS Yarrowia lipolytica (Candida lipolytica).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Dipodascaceae; Yarrowia.
OX NCBI_TaxID=4952;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=CLIB 122 / E 150;
RX PubMed=15229592; DOI=10.1038/nature02579;
RA Lafontaine I., de Montigny J., Marck C., Neuveglise C., Talla E.,
RA Goffard N., Frangeul L., Aigle M., Anthouard V., Babour A., Barbe V.,
RA Barnay S., Blanchin S., Beckerich J.-M., Beyne E., Bleykasten C.,
RA Boisarame A., Boyer J., Cattolico L., Confanioli F., de Daruvar A.,
RA Despons L., Fabre E., Fairhead C., Ferry-Dumazet H., Groppi A.,
RA Hantraye F., Hennequin C., Jauniaux N., Joyet P., Kachouri R.,
RA Kerrest A., Koszul R., Lemaire M., Lesur I., Ma L., Muller H.,
RA Nicaud J.-M., Nikolski M., Oztas S., Ozier-Kalogeropoulos O.,
RA Pellenz S., Potier S., Richard G.-F., Straub M.-L., Suleau A.,
RA Swennen D., Tekaiia F., Wesolowski-Louvel M., Westhof E., Wirth B.,
RA Zeniou-Meyer M., Zivanovic Y., Bolotin-Fukuhara M., Thierry A.,
RA Bouchier C., Caudron B., Scarpelli C., Gaillardin C., Weissenbach J.,
RA Wincker P., Souciet J.-L.;
RT "Genome evolution in yeasts.";
RL Nature 430:35-44 (2004).
DR EMBL; CR382128; CAG83062.1; -; Genomic DNA.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR011700; bZIP 2.
DR InterPro; IPR004827; TF_bZIP.
DR Pfam; PF07116; bZIP_2; 1.
DR SMART; SM00338; BRLZ; 1.
DR PROSITE; PS0217; bZIP; 1.
DR PROSITE; PS00036; bZIP_BASIC; UNKNOWN 1.
KW Complete proteome; DNA-binding; Nuclear protein.
SQ SEQUENCE 299 AA; 32817 MW; F526110CFB23ABB2 CRC64;

Query Match 10.0%; Score 230.5; DB 2; Length 299;
Best Local Similarity 29.7%; Pred. No. 0.00082;
Matches 90; Conservative 38; Mismatches 88; Indels 87; Gaps 13;

QY 18 ESFLSAPGDNFTSLPADSTPTLNPRDMTPDSVADIDSRSLVSPESQDAEDDESHSTA 77
DB 7 ESFTPTPDLSPLTADSPGS-----PESGD----- 32
QY 78 TAPSTSEKKPVKKRSGQVLPEKTNLPKRKATDEKEQRRVERVLNRRAAQSSRE 137
DB 33 -----KRRK-DLTLPLPAGALPPKRKATENEKEQRRIERIMNRQAHAARE 79
QY 138 KRRLEVEALEKR-----NKELETLLINVKTNLILVEE-----LNRFRSSGVVTRSS 186
DB 80 KRRHLEDEKKCSLSSENNDLHHQVTESKTNHLMHQHVSVAKLQQLSSLVNMAKS 139
QY 187 PLDSIQDSITLSQQLFGSRDGTMGNPEQSLMDQIMRGAANPTVNPASLPSLPISDK 246
DB 140 -----SGALAGV-DVPMDS--DVSMAPKLENPTAAPS-QPMGLA-SAPTLFNHD 183
QY 247 FQTKEEDE-----EQADEDEMEQWTHET-KEAAAAKKNKSKQSRVSTDS-----OR 293
DB 184 NETVVDPSPIVKTEEVDSNTFNLLHTESSPPPELAESTGSGSPSSLSLSCDETDLVDRAH 243
QY 294 PAV 296
DB 244 PAV 246
```



```

RESULT 11
HACL_YEAST
AC P41546; P87040; STANDARD; PRT; 230 AA.
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 13-SEP-2005 (Rel. 48, Last annotation update)
DE HACL1 protein.
GN Name=HACL1; Synonyms=ERN4, IRE2; OrderedLocusNames=YPL031W;
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
NCBI_TaxID=4932;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=S288c;
RA MEDLINE=95116316; PubMed=7816617;
RA Nojima H., Leem S.-H., Araki H., Sakai A., Nakashima N., Kanaoka Y.,
RA Ono Y.;
RT "HACL1: a novel yeast bZIP protein binding to the CRE motif is a
RT multicopy suppressor for cdc10 mutant of Schizosaccharomyces pombe."
RL Nucleic Acids Res. 22:5279-5288 (1994).
RN [2]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=S288c / AB972; PubMed=7670463;
RX MEDLINE=95400292;
RA Murakami Y., Naitou M., Hagiwara H., Shibata T., Ozawa M.,
RA Sasanuma S.-I., Sasanuma M., Teuchiya Y., Soeda E., Yokoyama K.,
RA Yamazaki M.-A., Tashiro H., Eki T.;
RT "Analysis of the nucleotide sequence of chromosome VI from
RT Saccharomyces cerevisiae."
RL Nat. Genet. 10:261-266 (1995).
RN [3]
RP SEQUENCE REVISION TO 183-230.
RA Murakami Y.;
RL Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.
RN [4]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=AB320;
RX MEDLINE=97222447; PubMed=9077435;
RA Mori K., Kawahara T., Yoshida H., Yanagi H., Yura T.;
RT "Signalling from endoplasmic reticulum to nucleus: transcription
RT factor with a basic-leucine zipper motif is required for the unfolded
RT protein-response pathway."
RL Genes Cells 1:803-817 (1996).
RN [5]
RP CHARACTERIZATION.
RX MEDLINE=97086687; PubMed=8932376; DOI=10.1093/nar/24.21.4222;
RA Nikawa J.-I., Akiyoshi M., Hirata S., Fukuda T.;
RT "Saccharomyces cerevisiae IRE2/HACL1 is involved in IRE1-mediated KAR2
RT expression."
RL Nucleic Acids Res. 24:4222-4226 (1996).
RN [6]
RP CHARACTERIZATION, AND ALTERNATIVE SPLICING.
RX MEDLINE=97053779; PubMed=8998193; DOI=10.1016/S0092-8674(00)81360-4;
RA Cox J.S., Walter P.;
RT "A novel mechanism for regulating activity of a transcription factor
RT that controls the unfolded protein response."
RL Cell 87:391-404 (1996).
CC -1- FUNCTION: Seems to be involved in the unfolded protein response
CC (UPR) pathway. Binds to the UPR element (UPRE) in the promoter of
CC UPR-regulated genes such as KAR2, PD11, EUG1 and PKB2; activates
CC the transcription of these genes.
CC -1- SUBUNIT: Homodimer.
CC -1- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Comment=the level of each isoform is regulated by a splicing
CC event that occurs when the UPR is induced by IRE1;
CC Name=U;
CC IsoId=P41546-1; Sequences=Displayed;
CC Name=I;

```

```

CC CC IsoId=P41546-2; Sequence=VSP_000586;
CC Note=Active and stable isoform which induces UPR;
CC -1- SIMILARITY: Belongs to the bZIP family.
CC -1- SIMILARITY: Contains 1 bZIP domain.
CC CC
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use as long as its content is in no way modified and this statement is not
CC removed.
CC CC
CC EMBL; D26506; BAA05513.1; -; Genomic DNA.
CC EMBL; D50617; BAA24425.1; -; Genomic DNA.
CC EMBL; D86413; BAA19565.1; -; Genomic DNA.
CC PIR; S78571; S78571.
CC GermOnline; 140124; -.
CC TRANSFAC; T02039; -.
CC Ensembl; YPL031W; Saccharomyces cerevisiae.
CC SGD; S00001863; HAC1.
CC GO; GO:0005634; C:nucleus; TAS.
CC GO; GO:0003704; F:specific RNA polymerase II transcription fa. .; TAS.
CC GO; GO:0003700; F:transcription factor activity; IDA.
CC GO; GO:0006644; P:phospholipid metabolism; TAS.
CC GO; GO:0045893; P:positive regulation of transcription, DNA-d. .; IDA.
CC GO; GO:0006357; P:regulation of transcription from RNA polym. .; TAS.
CC GO; GO:0006990; P:unfolded protein response, positive regulat. .; TAS.
CC InterPro; IPR011700; bZIP 2.
CC InterPro; IPR004827; TF_bZIP.
CC Pfam; PF07116; bZIP 2; 1.
CC SMART; SM00338; BRLZ; 1.
CC PROSITE; PS00217; bZIP; FALSE NEG.
CC PROSITE; PS00036; bZIP_BASIC; 1.
CC Activator; Alternative splicing; Complete proteome; DNA-binding;
KW Nuclear protein; Transcription; Transcription regulation;
KW Unfolded protein response.
FT DOMAIN 67 81 Leucine-zipper.
FT DNA_BIND 29 61 Basic motif.
FT VARSPLIC 221 230 AVITMTRKIQ -> EAQSLNSFELNDFPITS (in isoform I).
FT FTID=VSP_000586.
FT CONFLICT 143 230 RLCPRVCRFRVGRDFMGAAECLRRKMYQSRRLPYPTIN
FT NLFDAVASPLADPLCDDIAGNSLPDSDSLDLMWRNPVIT
FT MTRKIQ -> ATLSPKSMRDSASDQETSWELQMFKTENYPE
FT STTLPAVDNNLFDVAWRQWTHSATI (in Ref. 2).
FT RLCPRVCRFRVGRDFMGAAECLRRKMYQSRRLPYPTI
FT -> ATLSPKSMRDSASDQETSWELQMFKTENYPESTTLPAV
FT DN (in Ref. 4).
FT SQ SEQUENCE 230 AA; 26582 MW; 43073BCCCCC4709B CRC64;
Query Match 9.2%; Score 211; DB 1; Length 230;
Best Local Similarity 35.8%; Pred. No. 0.0055;
Matches 64; Conservative 26; Mismatches 51; Indels 38; Gaps 6;
QY 102 KTNLPKRKRAKTEDEKORRVERVLNRRAAQSSRRKRLEVEALEKRNKELETLL--IN 159
Db 23 KSTLPPKRKRAKTEKORRIIRNRRAHQSRKRLHLQYLERKSLLENLNSVN 82
QY 160 VQK-----TNLIIVELNFR-----RSSGVVTRSSSPSLDSLODSITLS----- 198
Db 83 LEKLADHEDALTCSDAFVASLDEYDFQSTRGASLDTRASS--HSSSDTFTPSPLNCTM 140
QY 199 -----OQLFGSRDQGMNPEQSLMDQIMRSAANPTVN---PASLSPSLPDISD 244
Db 141 EPLCRPRVCRFRVGRDFMGAAECLRRKMYQSRRLPYPTINLFDVASPLADPLCD 199
RESULT 12
Q6BQC2 DEBHA
ID Q6BQC2 DEBHA PRELIMINARY; PRT; 260 AA.
AC Q6BQC2;
DT 25-OCT-2004 (TREMBLrel. 28, Created)
DT 25-OCT-2004 (TREMBLrel. 28, Last sequence update)
DT 25-OCT-2004 (TREMBLrel. 28, Last annotation update)

```



```
RA Lafontaine I., de Montigny J., Marck C., Neveglise C., Talla E.,
RA Goffard N., Frangeul L., Aigle M., Anthouard V., Babour A., Barbe V.,
RA Barnay S., Blanchin S., Beckerich J.-M., Beyne E., Bleykaaten C.,
RA Boisrame A., Boyer J., Cattolico L., Confanioleri F., de Daruvar A.,
RA Despons L., Fabre E., Fairhead C., Ferry-Dumazet H., Groppi A.,
RA Hantraye F., Hennequin C., Jauniaux N., Joyet P., Kachouri R.,
RA Kerrest A., Kozul R., Lemaire M., Lesur I., Ma L., Muller H.,
RA Nicaud J.-M., Nikolski M., Oztas S., Ozier-Kalogeropoulos O.,
RA Pellenn S., Potier S., Richard G.-F., Straub M.-L., Suleau A.,
RA Swennen D., Tekoa F., Wesolowski-Louvel M., Westhof E., Wirth B.,
RA Zeniou-Meyer M., Zivanovic Y., Bolotin-Fukuhara M., Thierry A.,
RA Bouchier C., Caudron B., Scarpelli C., Gaillardin C., Weissenbach J.,
RA Wincker P., Souciet J.-L.,
RT "Genome evolution in yeasts.";
RL Nature 430:35-44(2004).
DR EMBL; CR380957; CAG61724.1; -; Genomic_DNA.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR011616; bZIP.1.
DR InterPro; IPR004827; TF_bZIP.
DR Pfam; PF00170; bZIP.1; 1.
DR SMART; SM00338; BRLZ; 1.
DR PROSITE; PS0217; bZIP; 1.
DR PROSITE; PS00036; bZIP_BASIC; 1.
DR Complete proteome; DNA-binding; Nuclear protein.
KW Complete proteome; DNA-binding; Nuclear protein.
SQ SEQUENCE 329 AA; 37162 MW; 469614BF3AF2282 CRC64;

Query Match 8.6%; Score 197; DB 2; Length 329;
Best Local Similarity 23.2%; Pred. No. 0.039;
Matches 66; Conservative 46; Mismatches 95; Indels 78; Gaps 6;

QY 102 KTNLPKRKAETDEKEQRRVRLNRRAAASSRERKLEVEALEKRNKELETLLINQV 161
DQ :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 26 ETWMPKRKAETKEEKEIRKIQLNRKAAQKSRDRKRYVYANLEKCKNTMKV---- 80

QY 162 KTNLLIVELNFRSSGVVTRSSSLDLSITLSQQLFGSRDQOTMSNPESQSLMDQI 221
DQ :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 81 -----LDQLQSKIDIKSMILDPFVVDVTYINMEQD--DEM 112

QY 222 MRSAAANTVNPASLSPSPSPKEFQTKKEEEDQADEDEMEQTHETKEAANAKEKNS 281
DQ :||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:
Db 113 SFSSDTPSPSTTEISSTLAENAVDVHAAEDVDNREYISSMIQ-----KKST 160

QY 282 KOSRVSTDSQTPAVSIGGDAAPVPSDDAGANCIGLDPVHDDGPFSGHSGFLSAALD 341
DQ :||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:
Db 161 KFGSIGSDVTSTPVRPRSTEQMTPLTASTSSSTC-----MSAYIS 200

QY 342 ADRYLLESQQLASPNASTVDDDYLAGDSAACTNPLPSDYDFDIN 386
DQ :||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:
Db 201 ASDNEVDKR-----QTITNTPLSASST-----PNKYDLRIN 232

RESULT 15
Q75B05 ASHGO
ID Q75EQ5_ASHCO PRELIMINARY; PRT; 228 AA.
AC Q75EQ5;
DT 05-JUL-2004 (TremBLrel. 27, Created)
DT 05-JUL-2004 (TremBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TremBLrel. 27, Last annotation update)
DE ACR216Cp.
GN Name=ACR216C;
OS Ashbya gossypii (Yeast) (Eremothecium gossypii).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Eremothecium.
OX NCBI_TaxID=33169;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=ATCC 10895;
RX PubMed=15001715; DOI=10.1126/science.1095781;
RA Dietrich F.S., Voegeli S., Brachat S., Lerch A., Gates K., Steiner S.,
RA Mohr C., Poehlmann R., Luedi P., Choi S., Wing R.A., Flavler A.,
RA Gaffney T.D., Philippsen P.
```

```
RT "The Ashbya gossypii genome as a tool for mapping the ancient
RT Saccharomyces cerevisiae genome.";
RL Science 304:304-307(2004).
DR EMBL; AB016816; AAS51442.1; -; Genomic_DNA.
DR AGD; ACR216C; -.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR011700; bZIP.2.
DR InterPro; IPR004827; TF_bZIP.
DR Pfam; PF07716; bZIP.2; 1.
DR PROSITE; PS00036; bZIP_BASIC; 1.
KW Complete proteome.
SQ SEQUENCE 228 AA; 24914 MW; 605A60D78633F038 CRC64;
```

```
Query Match 8.3%; Score 190; DB 2; Length 228;
Best Local Similarity 37.7%; Pred. No. 0.057;
Matches 49; Conservative 20; Mismatches 45; Indels 16; Gaps 3;

QY 102 KTNLPKRKAETDEKEQRRVRLNRRAAASSRERKLEVEALEKRNKELETLLINQV 161
DQ :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 14 KSTLPPKRKAETQEEKEQRRRIILNRKAAHQSRKKGLHLLYLERKCALLERIVAVD 73

QY 162 KTNLLIVE---ELNFRSSGVVTRSSSLDLSITLSQQLFGSRDQOTMSNPESQSLM 218
DQ :||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:
Db 74 LGALVAARGDAALGRAVTEYEAAREGSGTPSF--AITLDRSKNEG----- 120

QY 219 DQIMRSAANP 228
DQ :||:|:|:|
Db 121 PRVCRRTTP 130
```

Search completed: November 23, 2005, 03:24:56
Job time : 197.805 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 23, 2005, 03:07:01 ; Search time 31.6947 Seconds
(without alignments)
1369.117 Million cell updates/sec

Title: US-10-663-450-2
Perfect score: 2294
Sequence: 1 MAPQSSPLVKFEASPAESF.....QPQSGASHGCGDGGIAGV 451
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues
Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 80:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	242.5	10.6	238	2 S78571	transcription fact
2	170.5	7.4	688	2 T22750	hypothetical prote
3	154	6.7	962	2 T22459	hypothetical prote
4	153.5	6.7	2022	2 A59297	myosin X - mouse
5	153.5	6.7	5327	2 T13564	microtubule-associ
6	151	6.6	2484	2 T26216	hypothetical prote
7	151	6.6	2607	2 T26215	hypothetical prote
8	148.5	6.5	650	2 S16706	transcription acti
9	148.5	6.5	5170	2 T15348	hypothetical prote
10	147.5	6.4	2058	2 A59267	myosin X - human
11	146	6.4	1302	1 JC6009	surface-located me
12	145	6.3	2052	2 T18519	myosin X - bovine
13	144	6.3	1002	2 S70292	FUN12 protein - ye
14	143.5	6.3	168	2 T50922	bZIP protein HY5 l
15	143.5	6.3	326	2 T08591	TGAGG-motif bindin
16	143	6.2	406	2 S38170	SRP40 protein - ye
17	142.5	6.2	322	2 T08592	TGAGG-motif-bindin
18	141.5	6.2	1365	2 T30822	Imp1 protein - Myc
19	140.5	6.1	381	2 S26812	transcription fact
20	140.5	6.1	753	2 F96781	hypothetical prote
21	140	6.1	544	2 T40058	probable chromatin
22	140	6.1	1621	2 A82255	hypothetical prote
23	139	6.1	1132	2 T43483	translation initia
24	139	6.1	2722	2 T20532	hypothetical prote
25	138.5	6.0	1142	2 S59359	Gln4 protein - yea
26	138	6.0	619	2 T08613	hypothetical prote
27	138	6.0	624	2 PC6003	surface membrane p
28	138	6.0	768	2 H54024	protein kinase (EC
29	138	6.0	2526	2 T20531	hypothetical prote

RESULT 1

S78571
transcription factor HAC1 - yeast (Saccharomyces cerevisiae)
N:Alternate names: protein YFL031w
C:Species: Saccharomyces cerevisiae
C>Date: 13-Sep-1998 #sequence_revision 13-Sep-1998 #text_change 09-Jul-2004
C:Accession: S78571; S56223; S53578
R:Murakami, Y.
submitted to the Protein Sequence Database, January 1998
A:Reference number: S78570
A:Accession: S78571
A:Molecule type: DNA
A:Residues: 1-238 <MUR>
A:Cross-references: UNIPROT:P41546; UNIPARC:UPI0000157F61; EMBL:D50617; MIPS:YFL031w
A>Note: this is a revision to the sequence from reference S56186
R:Murakami, Y.; Naitou, M.; Hagiwara, H.; Shibata, T.; Ozawa, M.; Sasanuma, S.I.; Sasanuma
submitted to the EMBL Data Library, May 1995
A:Description: Analysis of the nucleotide sequence of chromosome VI from Saccharomyces cerevisiae
A:Reference number: S56186
A:Accession: S56223
A:Molecule type: DNA
A:Residues: 1-191, 'RWQTHSATI', <MUM>
A:Cross-references: UNIPARC:UPI0000179B60; EMBL:D50617; NID:G836685; PID:G836723; MIPS:YFL031w
A>Note: this sequence has been revised in reference S78570
A>Note: this was believed to be the complete sequence of YFL031w
R:Nojima, H.; Leem, S.H.; Araki, H.; Sakai, A.; Nakashima, N.; Kanaoka, Y.; Ono, Y.
Nucleic Acids Res. 22, 5279-5288, 1994
A:Title: Hac1: a novel yeast bZIP protein binding to the CRE motif is a multicopy suppressor of the
A:Reference number: S53578; MUID:95116316; PMID:7816617
A:Accession: S53578
A:Molecule type: DNA
A:Residues: 1-142, 'RLCRPRVCRPRVGRDFMGAAELRRKMYQSRRLVPTV', 183-220, 'AVITWTRKLQ' <NOJ>
A:Cross-references: UNIPARC:UPI000012C042; GB:D26506; NID:G633122; PIDN:BAA05513.1; PID:G633122
C:Genetics:
A:Gene: SGD:HAC1; IRE2; ERN4
A:Cross-references: SGD:S0001863; MIPS:YFL031w
A:Map position: 6L
A:Introns: 221/1
C:Keywords: DNA binding; nucleus; transcription factor

Query Match 10.6%; Score 242.5; DB 2; Length 238;
Best Local Similarity 31.4%; Pred. No. 8,7e-07;
Matches 76; Conservative 30; Mismatches 57; Indels 79; Gaps 9;
QY 102 KTNLPKRKAKTEDEKEQRRVERVLNRRRAAQSRRERKLEVEALEKRKNKELTLL--IN 159
Db 23 KSTLPKRKAKTKKEEQRRERILNRRRAAQSRRERKRLHLQYLERKCSLLENLNSVN 82
QY 160 VOK-----TNLILVEELNRRP-----RSSGVVTRSSSPDLSLQDSITLSQLFGS 204
Db 83 LEKLADHEDALTCSDHDAFVASLDEYDFQSTRGASLDTRASS-----124

protein F07A11.6 l
hypothetical prote
erythroid cell tra
hepatoma-derived g
protein kinase (EC
hypothetical prote
related to Ap-1-li
cyclin-dependent k
dextranase - Strept
microtubule-associ
hypothetical prote
ankyrin 2, neurona
hypothetical prote
hypothetical prote
hypothetical prote
hypothetical prote

QY 205 RDGQMSNPEQSLMDQIMR5AANPTVNPASLSPSPFISDKBFQTKEDDEQADEEME 264
Db 125 -----HSSDFTFTSPINCIMEPATLSP-----KSMRDSASDQE 158
QY 265 QWHTETKEAANAARAKSKSRVSTSTQRPAY-----SIGDDAAVPVFSDDAGANCLG 317
Db 159 TSW-----ELQMPKTEN-----VPSESTTLPAVDNNNLFDAVASPLADPLCDDIAG-NSLP 207
QY 318 LD 319
Db 208 FD 209
RESULT 2
T32750
hypothetical protein F57B10.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 09-Jul-2004
C:Accession: T32750
R:Greco, T.; Elliott, G.; Keppeler, D.
submitted to the EMBL Data Library, December 1997
A:Description: The sequence of C. elegans cosmid F57B10.
A:Reference number: Z21219
A:Accession: T32750
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-688 <GRE>
A:Cross-references: UNIPROT:O44743; UNIPARC:UPI0000079D06; EMBL:AF039713; PIDN:AAB96719.
A:Experimental source: strain Bristol N2; clone F57B10
C:Genetics:
A:Gene: CESP:F57B10.1
A:Map position: 1
A:Introns: 54/1; 150/3; 196/1; 306/3; 383/3; 627/3

Query Match 7.4%; Score 170.5; DB 2; Length 688;
Best Local Similarity 21.8%; Pred. No. 0.031;
Matches 94; Conservative 63; Mismatches 125; Indels 149; Gaps 17;
QY 4 QOSSPLVAFEAS-----PAESFLSAGDNFTSLFADSPFSLNPRDMWTPDSVADIDS 57
Db 187 QNSGLVRFKSKQPRVLPNPASISLNAPSSFNQSTSTPAT----- 228
QY 58 LSVIPESQDAEDDESHSTSAT-----APSTSEKK-----PVKKRKSQGVLPPE-- 100
Db 229 -----SSSSSTNGGFKVSGSTGERRKVPPLRLDEEIKLKCKEGICLPDFF 275
QY 101 PKTNLPKRKATEDEKEORRVERVLRNRRAAQSSRRKRLEVEAL-----KKNKEL 153
Db 276 PLTK-----AERDLKRIRKIRKRSQAOTSRKRQDYIEQLERDVSSESTKENQAL 326
QY 154 ETLINVKTNLILVLELNRRSSGVVTRSS-----PLDSLQDSITLSQOL--FGSR 205
Db 327 KQOIERLSENQSVISQLKQALQGNAKRTTQAGRCGLAVFMLSACLVLVSPQLSPLGNQ 386
QY 206 DQQTW-----SNPEQSLMD-----QIMRSA 225
Db 387 DNQKVLCTEEACQPSATSMNSANSQAIAIGVAPSVVIPSQGPVMVSTNANRQWNRNA 446
QY 226 A-----NPTVNPASLSPSPFISDKB-----FQTKEDDEQADEEMEQTWHTKEAA 274
Db 447 VLNHHNNSKYPASGNQNHHPHIALEDLNNHPPTLPQKSYQQO--HOPSMYRRSDTIAMA 504
QY 275 AAKENKSKSRVSTSTQRPAYSIGDDAAV-PVF-----SDAG-ANCL 316
Db 505 MAKIGARKGSTSSSSASVASSTSTSSATSPITRTLTGAFEDQCCDASSDDSCANNP 564
QY 317 GLDPVHQDDGP 327
Db 565 SLVPKMSAQ 575

RESULT 3
T22459

hypothetical protein F49E2.5b - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C:Accession: T22459
R:Sturston, J.
submitted to the EMBL Data Library, October 1994
A:Reference number: Z19566
A:Accession: T22459
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-962 <WIL>
A:Cross-references: UNIPROT:Q20625; UNIPARC:UPI000017B9BF; EMBL:Z46267; PIDN:CAA86429.1;
A:Experimental source: clone F49E2
C:Genetics:
A:Gene: CESP:F49E2.5b
A:Map position: X
A:Introns: 37/2; 66/3; 210/3; 284/3; 469/3; 499/3; 522/3; 585/3; 637/3; 779/1; 825/3; 866/3
Query Match 6.7%; Score 154; DB 2; Length 962;
Best Local Similarity 20.5%; Pred. No. 0.38;
Matches 88; Conservative 77; Mismatches 164; Indels 100; Gaps 18;
QY 59 SVIPESQDAEDDESHST--SATAPSTSEK-KPVKKRKSQGVLP-----PEPKTNLPKRKA 111
Db 41 AVLPSINDQRKEAFASLELTEQPOQVEKVKSEKKAQKQIAKHAEAEQKVN-----A 94
QY 112 KTEDEKQORRVERVLRNRRAAQSSRRKRLEVEAL----EKRNKELETLLINVKTNLIL 167
Db 95 KKAEEKARRAEAEAKKRAQEEHKKWKAQERIKQSEKKEADLKKLQAEKKKEKAVK 154
QY 168 VEELNRRSSGVVTRSSPLDSLQDSITLSQQLFGSRDGGQTMSPNEQSLMDQIMRSAAN 227
Db 155 AEKAEKAEK-----TKKASTPAPVEEIVVKV---ANDRSAAPAEPEK-----T 196
QY 228 PTVPASLSPSPFISDKBFQTKEDDEQO-----ADEDEMEQ---TWHTKEAANAARKE 278
Db 197 PTNPAPAEQVQOIEITGKKNKKNKSESEATAAPASVEQVQKPVVTEEPHQQAAPQE 256
QY 279 KNSKQSRVSTSTQRPAYSIGDDAAVPVFSDDAGANCLGDPVHODDGPFS----- 329
Db 257 KKNKKRKRKSEENVPAAS-----ETPV-----EPVVENESPESEPVAKLITV 299
QY 330 -----TCHSFGLSAALD-----ADRYLLESQSLASPNASTVDDDDYLADGSAACFTN 375
Db 300 SNTEASAVNVMGFSDIVTPKADEVITQDPVSAKQVLPVHVPSFIEPEEVA-----VS 352
QY 376 PLPSDYDFDINDFLT-----DDANHAAYDIVAASNYAAADRELDLEIHDPE---NQI-PS 426
Db 353 KKPTADSMDFLDFVTPKTEAESTSEAPAPVWVSKPTESIEDLEIVTYEHVADVGTGNTLSPS 412
QY 427 RHSIQQPOS 435
Db 413 QHSTPSPNS 421
RESULT 4
A59297
myosin X - mouse
C:Species: Mus musculus (house mouse)
C>Date: 09-Jun-2000 #sequence_revision 09-Jun-2000 #text_change 09-Jul-2004
C:Accession: A59297
R:Yonezawa, S.; Kimura, A.; Koshiba, S.; Masaki, S.; Ono, T.; Hanai, A.; Sonta, S.; Kageyama, R.; Yonezawa, S.; Kimura, R. Commun. 271, 526-533, 2000
Biochem. Biophys. Res. Commun. 271, 526-533, 2000
A:Title: Mouse myosin X: Molecular architecture and tissue expression as revealed by northern blot analysis
A:Reference number: A59297
A:Accession: A59297
A:Status: preliminary; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-2062 <YON>
A:Cross-references: UNIPROT:Q9JJY5; UNIPARC:UPI00000289A0; GB:AJ249706; PIDN:CAB56466.2
A:Experimental source: strain BALB/c; tissue lib testis
C:Genetics:
A:Gene: myo10

F;66-727/Domain: myosin motor domain homology <MMO>

Query Match	6.7%;	Score 153.5;	DB 2;	Length 2062;
Best Local Similarity	22.8%;	Pred. No. 1;		

Qy	91	RKSWGQVLPBPKTNLPBPKR-----AKTEDEKQORRVERVLNRRRAQSSRRKRLEVEA	145
Db	805	RVVYRQLLAB-KRELEBKREBEKKEEERERERARQREADLLURHQEATRRQOELEA	863
Qy	146	LEKRNKELETLLINVQKTNLILVELANFRSSGV--VTRSSSPLDLSIQSITLSLQQLFG	203
Db	864	LQKSORAAD-----LTRELEKQRENKQVEIILREKIEDLQ-----R	901
Qy	204	SRDQTMNSPEQSLMDQIMRASAANTVNPASLSPLPPISDKEFQTKEEDEEQADEDEM	263
Db	902	MKERQELSLTEASLQ-----KLQQLRDEELRLEDEACRAAQSFLE	942
Qy	264	BQWHTYKBAAAAKEKN-SKQSRVSTDSORPAVSIIGDAAVPVF-----	307
Db	943	SLAFDEIDECVRNTERISLVSIGSEBELGELAESASGEK--PSFNFSQPYPEEEVEDEG	1000
Qy	308	--SDDACANCLGLDPVHQDDGPPSIGHSGFLSAL-----DADRYLLESQLLASPNA-	357
Db	1001	FEADD-----DAFKDSNPNSBHGSHSDQRTSGIRTSDESSEEDPYMNDTVTPSPSD	1052
Qy	358	STV-----DDDYLI-----AGDSAACH-TNP--LPS---DYDFDINDFLTDDANHAAYDI	400
Db	1053	STVLLAASMQDSASLHNSSGESTYCNQPNQPGDLPSPDGDDYDQDDY-EDGALTSSGV	1111
Qy	401	VAASNYPAA	408
Db	1112	TFGNSYGS	1119

RESULT 5
Tl3564
microtubule-associated protein homolog - fruit fly (*Drosophila melanogaster*)
N;Alternate names: hypothetical protein EG:49E4.1
C;Species: *Drosophila melanogaster*
C;date: 13-Aug-1999 #sequence_revision 13-Aug-1999 #text_change 31-Dec-2004
C;Accession: Tl3564
R;Spanos, L.; Papagiannakis, G.; Siden-Kiamos, I.; Louis, C.
submitted to the EMBL Data Library, April 1999
A;Description: Sequencing the distal X chromosome of *Drosophila melanogaster*.
A;Reference number: Z17689
A;Accession: Tl3564
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-5327 <SPA>
A;Cross-references: UNIPROT:O76891; UNIPARC:UPI0000110149; EMBL:AL031128; PIDN:CAA20006
C;Genetics:
A;Cross-references: FlyBase:FBgn0025392
A;Introns: 24/2; 52/3; 104/3; 179/1; 232/1; 1669/3; 2566/1; 4798/3; 5272/1
A;Note: EG:49E4.1

Query Match	6.7%;	Score 153.5;	DB 2;	Length 5327;
Best Local Similarity	20.4%;	Pred. No. 3.2;		
Matches	98;	Conservative 84;	Mismatches 163;	Indels 135; Gaps 21;

Qy	68	EDDESHST	SATAPST	SEKPKVK	KKSGQVL	PEPKTNL	PPRKR	-----	AKT-EDEK	117		
Db	4140	EEDEHRFS	--PPSDVDKAI	-----	TPQMRPLSP	REEEVAKI	ADVAKV	KS	DK	4189		
Qy	118	-----	BQRVRLN	-----	RRAAOSSR	RKRKEVEA	-----	LEKRNKE	LETLI	158		
Db	4190	DITDI	IPDFBERQ	LEELK	STADTEES	DKSTRDEKS	LEISVK	VEISEK	SSPD	4243		
Qy	159	NVQKTNL	LILVELN	RRFRSSG	VTNR	-----	SSSP	--LDS	QDSTL	SQLFGSR	QGTMS	211
Db	4244	--QKSGP	ISIEEK	DKIK	EQSKA	QURQIG	LASSR	PESV	ASQPS	QSAAS	HEHKEVE	4301
Qy	212	NPQOSL	MDQIMR	--SAANPTV	NPASLS	PSLPIS	DKFEQ	TKEED	BEQAD	-----	258	

Db 4302 LSSSHKAEKSSRPESVASQSEKDMKTSRPASTSQFSTKEGDEETESLLHSLTTTETV 4361
 Qy 259 EDBEMEQ-----TWHTKEAAAAAEEKNSKQSRVST---DSTORPAVSI-- 298
 Db 4362 ETQMEKESFSESVSTSVTKSTVLSQSQTQLREESTSESLSSSLKVEDSSRRESLSLL 4421
 Qy 299 ---GGDANVPVSDDGAGNCLGLDPV---HQDGPFSIGHSGLSAA-----LOADRY 345
 Db 4422 AERGGIATNTSLKEDTSSASASQLEELLVQSECSSESIVSEIQTSIAQKSNKEIKDARET 4481
 Qy 346 LLESQLLASPNASTVDDDYLAGDSAACTNPLPSDYDFDINDFLTDDANHAAYDIV--- 401
 Db 4482 KVTSQFTTTTSSAYKODS--LKETVAEF---LATEKIVSKAEAFSTEATKSAODCLKXTT 4536
 Qy 402 --AASNYAAADREL-----DLEIHDPENQIPSRHSIQPOQS 435
 Db 4537 ASAVSSTSASQALFVCTDRESRESLLSQASESRLTHSDPEDEPADDVDRSSVKESRS 4596
 RESULT 6
 T26216
 hypothetical protein W06A7.3c - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
 C:Accession: T26216
 R:Ainscough, R.
 submitted to the EMBL Data Library, August 1996
 A:Reference number: Z20173
 A:Accession: T26216
 A:Status: preliminary; translated from GB/EMBL/DBSJ
 A:Molecule type: DNA
 A:Residues: 1-2484 xWIL>
 A:Cross-references: UNIPROT:Q9U347; UNIPARC:UPI00000827E9; EMBL:Z78066; PIDN:CA47
 A:Experimental source: clone W06A7
 C:Genetics:
 A:Gene: CESP:W06A7.3c
 A:Map position: 5
 A:Introns: 1827/1; 1866/3; 1963/3; 1990/3; 2262/1; 2289/1; 2339/2; 2463/2
 Query Match 6.6%; Score 151; DB 2; Length 2484;
 Best Local Similarity 21.3%; Pred. No. 1.8; Indels 156; Gaps 23;
 Matches 111; Conservative 66; Mismatches 66
 Qy 37 PSTLNPRDMTPDSVADIDSRLSVIP-----ESQDAEDDESHSTSATAFSTSEKK 86
 Db 1644 PKYAKPLD-----DSKSRVTFAPINIKLGITYSEEQKELVESLERPLTI-ITQKP 1694
 Qy 87 PVKGRKSWGQVLPEPKTNLPPRKRAKED-----EKEQRRVERVURNRAAQ 133
 Db 1695 PERPTEDIGALSP-----LSPNTLAEYEEVPMMDMQSPHSPQEKQBEIEALSIIIEEPQ 1749
 Qy 134 SSERKRELEVALEKRNKELETLLINVQTNLTIVE-----ELNRRF----- 175
 Db 1750 AMKEVEKPVESAPEKONESLEAPEINIEPIRRVLVETKIMGPKSLNEDNDDDDGSECL 1809
 Qy 176 ----RSGGVVTRSSSPLD--SLQDSITLSLQQLFGSRD--GQTMNSN-----PEQSLMDQ 220
 Db 1810 DSTGDLSERTIQFNTSIDDPISRRDSFSSISFGGRQKFRTAENIRQDLLPFQSSVSQ 1869
 Qy 221 IMRSAANPTV-----NPASLSPSLPPI---SDKEFQTK--BEDEEQAD---EDEE 262
 Db 1870 YLRSSPNPSQQLLVTLNMDSPSDLSFNAPVPVFENTAQLFLEKLQCEDRPSAEGSIDSSG 1929
 Qy 263 MEQWTWHTKEAAAAAEEKNSKQSRVSTDSQRPVAVSICGGDAAVVPFSDDAGANCLGLDPVH 322
 Db 1930 FEKVDHEGLDEFAAPVPH-----DPMQK---SVFGSLG-----SDD-----MKPGS 1967
 Qy 323 QDDGPFSGHSGFLSALDADRYLLESQLLASPNASTVDDDYLAGD--GAACFTNPLPS- 379
 Db 1968 QDDGPFVIERNEANEATLK-----KNQKMSHHNDVIEKNYFNNDNAPTAALESIAEE 2021
 Qy 380 -----DYDFDINDFLT-----DANHAAYDIVAASN 405

Db	2022	ARKLVQDAVESAYEYKQAVDSGEIGRELLDNVEQKIBQVKEPIVDSLHKAYDVGVDV	2081
Qy	406	YAAARELDLEIHPENQIPSR---HSIQOQSGASSHCDD	444
Db	2082	HETVPNAVDDFVREAEKQLPESVPKEIETPPLVDIHTVD	2123
RESULT 7			
T26215			
hypothetical protein W06A7.3a - Caenorhabditis elegans			
C:Species: Caenorhabditis elegans			
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004			
C:Accession: T26215			
R:Ainscough, R.			
submitted to the EMBL Data Library, August 1996			
A:Reference number: Z20173			
A:Accession: T26215			
A>Status: preliminary; translated from GB/EMBL/DBJ			
A:Molecule type: DNA			
A:Residues: 1-2607 <WIL>			
A:Cross-references: UNIPROT:Q23187; UNIPARC:UPI000007623E; EMBL:Z78066; PIDN:CAB01522.2;			
A:Experimental source: clone W06A7			
C:Genetics:			
A:Gene: CESP:W06A7.3a			
A:Map position: 5			
A:Introns: 1827/1; 1866/3; 1963/3; 1990/3; 2262/1; 2289/1; 2412/1; 2462/2; 2586/2			
Query Match			
Best Local Similarity 6.6%; Score 151; DB 2; Length 2607;			
Matches 111; Conservative 66; Mismatches 189; Indels 156; Gaps 23;			
Qy	37	PSTLPRDMMTPDSVADIDSLSVIP-----ESODAEDESHSATSATPSTSEKK	86
Db	1644	PKVAPLD-----DSKSRVFAPLNLKGRTYSEEQKELVSLERPLTI-ITQKP	1694
Qy	87	PVKGRKSGQVLPPEKTNLPKRKAKTED-----EKEQRRVERVLRNRAAQ	133
Db	1695	PEKPTEDIGALSP----LSPNTLAERYEVPWMDQSVPHSQEKQEEIEALSEIEEPQ	1749
Qy	134	SSRERKRLEVEALEKRNKELELLINQVKTNILVE-----ELNRFR-----	175
Db	1750	AMKEVEKPVESAPEKDNESLEAPEINIEPIRVLVETKIMGPKSLNEDNDDDDGSECL	1809
Qy	176	-----RSSGVFRSSPLD--SLQDSITLSQQLFSGRD--GQTMGN-----PEQSLMDQ	220
Db	1810	DSIGDLSERTIQRFNTSIDPIRDSFSSISFGDRQKRFRTAENIRQDLPLFQSSVSQ	1869
Qy	221	IMRSAANPTV-----NPAISLSPSLPPI---SDKEFQTK---EDEEQAD---EDEE	262
Db	1870	YLRSNPNPSQQLLVNLSMDSFSLSPNAPPVGFENTAQFLEKLQEDRPSAEGSIDSSG	1929
Qy	263	MQQTHETKEAATAAKENKSKQSRVSTDSQRPVAVSIGGDAAVPVFSDDGAGANCLGLDPVH	322
Db	1930	FEKVDHEGLDEFAAPPVH-----DPMQK---SVFGSLG----SDD-----MKFGS	1967
Qy	323	QDDGPFISGHSGLSAALDADRYLLESQLLASPNASTVDDDLAGD--SAACFTNPLPS-	379
Db	1968	QDDGFVFIERNANEATLK-----KNQKSGSHNDVIEKNFTNDNAPTALESPIABE	2021
Qy	380	-----DYDFDINDFLTD-----DANHAAYDIVAASN	405
Db	2022	ARKLVQDAVESAYEYKQAVDSGEIGRELLDNVEQKIBQVKEPIVDSLHKAYDVGVDV	2081
Qy	406	YAAARELDLEIHPENQIPSR---HSIQOQSGASSHCDD	444
Db	2082	HETVPNAVDDFVREAEKQLPESVPKEIETPPLVDIHTVD	2123
RESULT 8			
S16706			
transcription activator YAP1 - yeast (Saccharomyces cerevisiae)			
N:Alternate names: hyper-resistance-conferring protein SNQ3; PAR1 protein; protein YM957			

C:Species: Saccharomyces cerevisiae			
C>Date: 28-May-1993 #sequence_revision 28-May-1993 #text_change 09-Jul-2004			
C:Accession: S16706; S19639; S19639; A30165; S48512; JE0416; S55112; S14832; S17002			
R:Hertle, K.; Haase, E.; Brendel, M.			
Curr. Genet. 19, 429-433, 1991			
A>Title: The SNQ3 gene of Saccharomyces cerevisiae confers hyper-resistance to several f			
A:Reference number: S16706; MUID:91347405; PMID:1878996			
A:Accession: S16706			
A:Molecule type: DNA			
A:Residues: 1-650 <HER>			
A:Cross-references: UNIPROT:P19880; UNIPARC:UPI0000052EC3; EMBL:X63268; NID:G4505; PIDN:			
R:Schneil, N.; Entian, K.D.			
Eur. J. Biochem. 200, 487-493, 1991			
A>Title: Identification and characterization of a Saccharomyces cerevisiae gene (PAR1) c			
A:Reference number: S19639; MUID:91364700; PMID:1889413			
A:Accession: S19639			
A>Status: nucleic acid sequence not shown; not compared with conceptual translation			
A:Molecule type: DNA			
A:Residues: 1-650 <SCH>			
A:Cross-references: UNIPARC:UPI0000052EC3			
R:Moyle-Rowley, W.S.; Hareham, K.D.; Parker, C.S.			
Genes Dev. 3, 283-292, 1989			
A>Title: Yeast YAP1 encodes a novel form of the jun family of transcriptional activator			
A:Reference number: A30165; MUID:89252807; PMID:2542125			
A:Accession: A30165			
A:Molecule type: DNA			
A:Residues: 1-585 'E', 587-650 <MOY>			
A:Cross-references: UNIPARC:UPI0000168EFE; GB:X58693; NID:G4797; PIDN:CAA41536.1; PID:G4:			
R:Schneil, N.			
submitted to the EMBL Data Library, June 1991			
A:Reference number: S48512			
A:Accession: S48512			
A:Molecule type: DNA			
A:Residues: 1-315 'S', 317-650 <SCH2>			
A:Cross-references: UNIPARC:UPI0000069D64; EMBL:X60780; NID:G499192; PID:G499193			
R:Husain, M.; Lenard, J.			
Gene 101, 149-152, 1991			
A>Title: Characterization of PDR4, a Saccharomyces cerevisiae gene that confers pleiotro			
A:Reference number: JE0416; MUID:91285426; PMID:2060792			
A:Accession: JE0416			
A:Molecule type: DNA			
A:Residues: 1-290, 'L'KPNLNMKFRSLVR', 305-647, 'D', 649-650 <HUS>			
A:Cross-references: UNIPARC:UPI0000168D62; GB:X53830; NID:G4121; PIDN:CAA37827.1; PID:G4:			
R:Gentles, S.; Bowman, S.			
submitted to the EMBL Data Library, June 1995			
A:Reference number: S55102			
A:Accession: S55112			
A:Molecule type: DNA			
A:Residues: 1-650 <GEN>			
A:Cross-references: UNIPARC:UPI0000052EC3; EMBL:Z49810; NID:G854472; PID:G854483; MIPS:Y:			
A:Experimental source: strain AB972			
C:Genetics:			
A:Gene: SGD:YAP1; PDR4; SNQ3; PAR1			
A:Cross-references: SGD:S0004466; MIPS:YML007W			
A:Map position: 13L			
C:Function:			
A>Description: transcription activator			
A>Note: confers pleiotropic drug resistance when present in high copy number			
C:Keywords: antibiotic resistance; DNA binding; leucine zipper; nucleus; transcription r			
F:59-99/Domain: fos/jun DNA-binding domain homology <FUD>			
F:63-156/Domain: DNA binding #status predicted <DNB>			
F:92-120/Region: leucine zipper motif			
Query Match			
Best Local Similarity 6.5%; Score 148.5; DB 2; Length 650;			
Matches 106; Conservative 85; Mismatches 171; Indels 239; Gaps 23;			
Qy	44	DMWTPDSVADIDSLRSLVPIESQDAEDDESHSATSATPSTSEKKPVKGRKSGQVLPPEKT	103
Db	11	DVSPGSLAEFEG-----SKSEHDEIENEHRTTGTGDEGDSQPKKGS-----	54
Qy	104	NLPPPKRAKTEDEKQRRVERVLRNRAAQSSRRKRLEVEALEKRNKELETLINQVKT	163


```
Db 55 -----KTSKKQDLDPETKQKRTAQNRAAQAQAFRERKQKLEKVKQSL-----SIQQ 105
QY 164 N-----LILVELNRFR-----RSSGVVTRSSPLDS 190
Db 106 NEVEATFLRLDQLITLNLKLYRPTETRNDSKVLEYLARRDPNLHFSKNVNHNSSEPIDT 165
QY 191 LODSI-----TSLQQLFGRDQGTMSNPEQSL-----MQQIRSAAN 227
Db 166 PNDIDQENVKQKNFTFQYPLDNDNDNSKNVGKQLPSPNDFSHAPINQTKKUSD 225
QY 228 PTVPNPSLSPSPISD-----KEFQTKKEDEEQA 257
Db 226 ATDSSSATLDSLNSNDVNLNTPNSSTMDLNDVLYTNRFVSGDDGSKNTKLDNSMF 285
QY 258 DDEEMEQTWHE-----TKB-----AAAAKEKSKQSRVSTDSQTPA 295
Db 286 SDFNFENQFDEQVSEFCSKMNQVCGTRQCPKPKPISALDKVEFASSTLSNS--PA 342
QY 296 VSIIGDAAVPFSDAGANCLGDPVHODDGPFSIGHSGF--LSAALDADRYLLESQLLAS 354
Db 343 LNTWESHNI--TDNTPANVIATDATKYENS--FS--GFRGLGFDMSAHNVVNDNSGTS 397
QY 355 PNA-----STVDDDDL-----AGDSAAACFTNP----- 376
Db 398 TDSGTGCKNKKNNNSDDVLPFISESPDMQVNTNPFSPGSGTGIGNNAASNTNPSLLQ 457
QY 377 -----LPSD-----YDFDI-----NDFLTLD 391
Db 458 SSKEDIPFINANLAFDDNSTNIQLQPFSESQNKFDYDMFDRDSSKEGNNLFGFLED 517
QY 392 DANHAAYDIVASNAADREL---DLEIHDPE-----NQIPSRHSIQPOSGASSHCC 442
Db 518 DB-----DDKKAANMSDDESSLIKQLINEBPELPKQYLQSVPGNESEISQKXGSLQNA 572
QY 443 D 443
Db 573 D 573

RESULT 9
T15348
hypothetical protein B0350.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 09-Jul-2004
C:Accession: T15348
R:Gattung, S.
A:Submitted to the EMBL Data Library, February 1996
A:Description: The sequence of C. elegans cosmid B0350.
A:Reference number: Z18332
A:Accession: T15348
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-5170 <GAT>
A:Cross-references: UNIPROT:Q17450; UNIPARC:UPI0000168675; EMBL:U50071; NID:g1208871; PI
A:Gene: CESP:B0350.1
A:Introns: 48/1; 5039/3; 5116/3

Query Match 6.5%; Score 148.5; DB 2; Length 5170;
Best Local Similarity 19.2%; Pred. No. 5.9;
Matches 96; Conservative 76; Mismatches 203; Indels 125; Gaps 19;

QY 16 PAESFLSAGDNFTSLFADSTPST-----LNPRDMTTPDSVAIDSLRSLVIP----- 62
Db 3757 PEDSHSPVSEDVHGFKVTKITTTTTTTHHEPEPDHTSDEHVESERYASGSPVPSSEN 3816
QY 63 -----ESQDAEDESHSTSATAPSTSEKKPVKKRKSQVLPPEKTNLPP 107
Db 3817 SNRVTTTTTTTTRHEPEPDQEHVVESQSYAS--GSPVSEKSVKVIETTTTTT 3875
QY 108 RKRATEDEKEQRRVRLNRRAAQSRKRKRLEVEALEKKNKELE--TLLINVOKTNL 165
Db 3876 REHFHEDE-----IPTIVESHDDQAAS-----SVPSEEDVHGQIQITTTTTT 3925
```

```
QY 166 ILVBEINFRSRSSGVVTRSSPLDSQDSITLSQ-----QLFGSRD 206
Db 3926 VPDEIDSGRMDLEKYSSESPPPEEDSSRVIETTTTTTFIREHPEPDDHSHVVGSOE 3985
QY 207 QOTMSN--PEQSLMDQIMRSAANPTVNPASLSPSPPIISDKEFQTKEE-----DEEQ 256
Db 3986 YSAGSPVPSEKSVRVERVIETTTTT-----TSEHFEHEDEHAQLHDTVSSBK 4033
QY 257 ADEDEMEQWTHETKEAAAAAKEKSKQSRVSTDSQTPAIVSIGDAAVPFSDDAGANCL 316
Db 4034 TSSSEPVQTLLEADITETHQRFSAEDEHQPTSTKPV-----HCFMETTSTSHASHV 4088
QY 317 GLDPVHQDGPFSIGHSGF--LSAALDADRYLLESQLLASPNASTVDDYLAGSAAACF 373
Db 4089 ELE--HRDDDDSGEGGFGSKVLGPAKKAG--MVAGGVVAAPVALAA-----VGAKAA-- 4137
QY 374 TNPLPSDYDFDINDPLTDDANHAAYDIVAASNYAAADRE---LDEIHDPEHQIPSRHS 429
Db 4138 -----YD-----ALKODDDEATYDL-----DREHGSYESAELQRPQQVESROE 4177
QY 430 IQPOSGASSHGCDGDIIV 449
Db 4178 DSESGSPVDHGDSQHLPV 4197

RESULT 10
A59267
myosin X - human
C:Species: Homo sapiens (man)
C:Date: 02-Jun-2000 #sequence_revision 02-Jun-2000 #text_change 09-Jul-2004
C:Accession: A59267
R:Rogers, M.S.; Strehler, E.E.
A:Submitted to GenBank, March 2000
A:Description: Identification of myosin X as a specific binding partner for the tumor ser
A:Reference number: A59267
A:Accession: A59267
A>Status: preliminary; not compared with conceptual translation
A:Molecule type: mRNA
A:Residues: 1-2058 <ROG>
A:Cross-references: UNIPROT:Q9HD67; UNIPARC:UPI0000042862; GB:AF234532; PIDN:AAF37875.1
A:Gene: MYO10
F:66-727/Domain: myosin motor domain homology <MMO>

Query Match 6.4%; Score 147.5; DB 2; Length 2058;
Best Local Similarity 23.5%; Pred. No. 2.2;
Matches 86; Conservative 58; Mismatches 119; Indels 103; Gaps 18;

QY 91 RKSNGQVLPEPKTNLPPRK-----RAKTEDEKEQRRVRLNRRAAQSRKRKRLEVEA 145
Db 805 RRVYQLLAEREQBEKKQBEKKKEBEERERERERERERERERERERERERERERER 864
QY 146 LEKRNKELE--TLLINVOKTNLILVEELNFRSSGVVTRSSPLDSQDSITLSQQLFGS 204
Db 865 LQKSQKEAELTRELEKQKKNK--QVEEILRLK-----EIEDLQRMKEQ----- 906
QY 205 RDGQTMSPQSLMDQIMRSAANPTVNPASLSPSPPIISDKEFQTKKEDEQA----- 257
Db 907 ---QELSITIASLQ-----KLQERRDQELRLRLEEACRAAQEFLES 944
QY 258 ---DEDEB-----MEQWTHETKEAAAAAKEKSKQSRVSTDSQ---RPAVSTGGDAVPVF 307
Db 945 LNFDEICVNRNIRSISVGSEFSLEAESACEKPNFNSQPYPEEVEDEGFEADDAF 1004
QY 308 SDDAGANCLGDPVHQDGPFSIGHSGF--FGLSAALDA---DRYLLSOLLASPN--ST 359
Db 1005 KD-----SPNPEHGHSDQRTSGIRTSDDSEEDPYMNDTVPTSPSALST 1050
QY 360 V-----DDLYL-----AGDSAAACF---TNPLPS---DYDFDINDPLTDDANHAAYDIVA 402
Db 1051 VLLAPSVQDSSGLNSHSSGSESTYCMQPQAGDLPSPDGDYDYDQDDY--EDGAIITSGSVTF 1109
```


A:Molecule type: DNA
A:Residues: 'MLK', 134, 'LLRLRLRLRNPKNL', 149, 'VKVINQVILRSQPR', 165, 'YCRGCFETS', 177-180, 'NLSR', 472-640, 'Q', 642-721, 'H', 723-969, 'I', 971-1000, 'SNERGIR' <SUT>
A:Cross-references: UNIPARC:UPI0000168C1B; EMBL:L29389; NID:G460043; PIDN:AAA57228.1; PIR:Q039895
C:Genetics:
A:Gene: SGD:FUN12
A:Cross-references: SGD:S0000033; MIPS:YAL035W
A:Map position: 1L
C:Keywords: GTP binding; nucleotide binding; P-loop; transmembrane protein
F:406-533/Domain: translation elongation factor Tu homology <ETU>
F:412-419/Region: nucleotide-binding motif A (P-loop)
F:491-507/Domain: transmembrane #status predicted <TM1>
F:530-533/Region: GTP-binding NKXD motif

Query Match 6.3%; Score 144; DB 2; Length 1002;
Best Local Similarity 24.0%; Pred. No. 1.4;
Matches 60; Conservative 41; Mismatches 69; Indels 80; Gaps 10;

QY 62 PESQDAEDDESHSTSATPST-----SEKPKVKRKSQVLPPEKTNLPP 108
DB 35 PESSAGADTTSREASASAGAEIEGDFMSTLKQSKKQKQK-----VIEKKDQKPKIL 88
QY 109 KRAKTEDEKEQRRVERVLNRRAAQSRRERKLEVEALEKRNKELETLLINQVKTNLILV 168
DB 89 K-SKKEKEKEKEKQKKKEQAA-----RKAQQAQKQKQKEL-----NKQVKEKAAA 137
QY 169 EE--LNRFRSSGVVTRSSSP-----LDSLQDSITISQQLFGSRDQGTWSNPPQ 215
DB 138 EKAAAEKSKQSGESKPSASAKPAKVPAGLAALRQLELKKQL-----EQ 186
QY 216 SLMDQIMRSAAANTVNPASLSPLPSIDKEFTQKDEEEOADEEMEQTWHTEKAAA 275
DB 187 EKLER-----BEEERLEKEEERLANEKKMK-----EAKAAKK 220
QY 276 AKEKNSQSR 285
DB 221 EKEKAKREKR 230

RESULT 14
TS0922
bZIP protein HV5 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 21-Jul-2000 #sequence_revision 21-Jul-2000 #text_change 09-Jul-2004
A:Accession: TS0922
R:Oyama, T.; Shimura, Y.; Okada, K.
A:Description: The Arabidopsis HV5 gene encodes a bZIP protein that regulates stimulus-i
A:Reference number: Z25271
A:Accession: TS0922
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-168 <OYA>
A:Cross-references: UNIPROT:Q24646; UNIPARC:UPI000012CF4; EMBL:AB005295; PIDN:BA211116
A:Experimental source: Landsberg erecta

Query Match 6.3%; Score 143.5; DB 2; Length 168;
Best Local Similarity 30.5%; Pred. No. 0.18;
Matches 47; Conservative 19; Mismatches 57; Indels 31; Gaps 4;

QY 29 TSLFADSTPTLNPRDMTP-----DSVADIDSRLSVIP-----ESQDAEDDESHSTSATAP 80
DB 7 SSLAASSLPSSERSSSSAPHLEIKEGIESDEIRRVPEFGGAVGKTSRGSGSAT-- 64
QY 81 STSEKKPKVKRKSQVLPPEKTNLPFRKRAKTEDEKEQRRVERVLNRRAAQSRRERK 140
DB 65 -----CQRTQATVGESQRRKGRTPAEKENKRLKLLRNRVSAQAQARERK 110
QY 141 L-----EVEALEKRNKELETLLINQVKTNLIL 167
DB 111 AYLSELENRVKDLNKENSELERLTLQENQML 144

RESULT 15

T08591
TGACG-motif binding protein STF1 - soybean
C:Species: Glycine max (soybean)
C:Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 31-Dec-2004
A:Accession: T08591
R:Cheong, Y.H.; Yoo, C.M.; Park, J.M.; Ryu, G.R.; Goekjian, V.H.; Nagao, R.T.; Key, J.L.;
submitted to the EMBL Data Library, September 1995
A:Description: STF1 is a novel TGACG-binding factor with a zinc-finger motif and a bZIP
A:Reference number: Z16445
A:Accession: T08591
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-326 <CHB>
A:Cross-references: UNIPROT:Q39895; UNIPARC:UPI00000A6233; EMBL:L28003; NID:G2934883; PIR:
A:Experimental source: strain Williams; hypocotyl
C:Superfamily: TGACG-motif-binding transcription factor

Query Match 6.3%; Score 143.5; DB 2; Length 326;
Best Local Similarity 29.1%; Pred. No. 0.4;
Matches 43; Conservative 27; Mismatches 51; Indels 27; Gaps 5;

QY 48 PDSVADIDSRLSVIPESQDAEDDESHSTSATPSTSEKKPKVKRKSQVLPPEKTNLPP 107
DB 195 PES-----DEEIRRVPEI-----GGESAGTSASRPDAGSNAGTERAQTGD-----SQ 237
QY 108 RKRKTEDEKEQRRVERVLNRRAAQSRRERK-----LEVEALEKRNKELETLLINV 160
DB 238 KKRGRSPADKESKRLKRLLRNVSAQAQARERKKAYLIDLETRVKDLEKKNSELKERLSTL 297
QY 161 QKTNLIL-----VEELNFRFRSSGVVTRSS 185
DB 298 QNENQMLRQILKNTTASRRSGNSGTNNA 325

Search completed: November 23, 2005, 03:26:02
Job time : 34.6947 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 23, 2005, 03:04:00 ; Search time 176.139 Seconds
(without alignments)
1125.017 Million cell updates/sec

Title: US-10-663-450-2
Perfect score: 2294
Sequence: 1 MAFQSSPLVKFEASPSF.....QPQSGASHGCGDGGIANGV 451

Scoring table: BLOSUMP62
Gapop 10.0 , Gapext 0.5

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_21.*
1: Geneseqp1980s:*
2: Geneseqp1990s:*
3: Geneseqp2000s:*
4: Geneseqp2001s:*
5: Geneseqp2002s:*
6: Geneseqp2003as:*
7: Geneseqp2003bs:*
8: Geneseqp2004s:*
9: Geneseqp2005s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2294	100.0	451	5 AAE15371	Trichoder
2	2294	100.0	451	8 ADS12801	Trichoder
3	2294	100.0	451	8 ADS12804	Trichoder
4	2279.5	99.4	450	4 AAB82975	Trichoder
5	437.5	19.1	350	4 AAB82976	Aspergill
6	437.5	19.1	350	5 AAE15372	Aspergill
7	434.5	18.9	386	5 AAE15379	Aspergill
8	434.5	18.9	386	8 ADS12815	Aspergill
9	434	18.9	349	8 ADS12805	Aspergill
10	434	18.9	349	8 ADS12803	Aspergill
11	434	18.9	349	8 ADS12863	Aspergill
12	429.5	18.7	409	7 ABO43144	A. thalia
13	429.5	18.7	409	7 ADB31925	Plant (A.
14	429.5	18.7	409	8 ADO02271	Thalacro
15	426.5	18.6	342	4 AAB82977	Aspergill
16	426.5	18.6	342	5 AAE15381	Aspergill
17	426.5	18.6	342	8 ADS12818	Aspergill
18	323	14.1	64	5 AAE15373	Trichoder
19	288	12.6	64	5 AAE15374	Aspergill
20	242.5	10.6	230	2 AAW53806	transcrip
21	242.5	10.6	230	8 ADT87049	Yeast Str
22	242.5	10.6	238	2 AAW53807	transcrip
23	231.5	10.1	200	8 ADS43437	Bacterial
24	196	8.5	68	5 AAE15382	Yeast HAC

25	196	8.5	68	8	ADS12859	Saccharom
26	196	8.5	84	5	ABP02534	Human ORF
27	166	7.2	3257	4	ABB67502	Drosophili
28	165.5	7.2	192	4	AAB82615	Maize taa
29	165.5	7.2	192	4	AAB82616	Maize roo
30	165.5	7.2	192	4	AAG66526	Maize roo
31	163.5	7.1	1442	4	ABG25601	Novel hum
32	163.5	7.1	2408	4	ABG10631	Novel hum
33	163	7.1	192	4	AAB82614	Maize roo
34	163	7.1	192	4	AAG66525	Maize roo
35	148.5	6.5	650	6	ABR52838	Ptcutin s
36	148.5	6.5	650	7	ADK62274	Diapara
37	148.5	6.5	869	3	AAB42020	Human ORF
38	148.5	6.5	2048	4	AAE11891	Angiogene
39	148.5	6.5	2057	4	AAE11890	Angiogene
40	147.5	6.4	2058	5	ABB97219	Novel hum
41	147.5	6.4	2058	6	ABP71989	Human myo
42	147.5	6.4	2058	6	ABU03558	Angiogene
43	147.5	6.4	2058	7	ABO84401	Human myo
44	147.5	6.4	2058	8	ADL61264	Human pro
45	147.5	6.4	2058	8	ADQ17925	Human sof

ALIGNMENTS

RESULT 1
AAE15371
ID AAE15371 standard; protein; 451 AA.

XX AAE15371;

AC AAE15371;

XX 29-AUG-2003 (revised)

DT 07-MAR-2002 (first entry)

XX Trichoderma reesei HAC1 protein.

XX Heterologous protein secretion; unfolded protein response; UPB; lipase;

KW cellulase; carbohydrase; industry; purification; HAC1 protein.

XX Hypocrea jecorina.

XX Key Location/Qualifiers

FT Binding-site 84..147

FT /label= DNA-binding_domain

XX US2001034045-A1.

XX 25-OCT-2001.

XX 23-MAR-2001; 2001US-00816277.

XX 24-MAR-2000; 2000US-00534692.

XX (GENV) GENENCOR INT INC.

XX Penttila MB, Ward M, Wang H, Valkonen MJ, Saloheimo MJA;

XX WPI: 2002-033728/04.

XX N-PSDB; AAD24595.

XX Increasing secretion of heterologous proteins e.g. lipase and cellulase

PT in eukaryotic cells useful in industry to increase production and

PT facilitate purification, by inducing an elevated unfolded protein

XX Claim 39; Fig 7; 56pp; English.

XX The present invention relates to methods for increasing the secretion of

CC heterologous protein in eukaryotic cells by inducing an elevated unfolded

CC protein response (UPR). The method involves inducing the elevated UPR by

CC increasing the presence of proteins such as HAC1, HACA, PTC2 or IRE1 in

CC cells. The method and sequences are useful for increasing the secretion

CC	of heterologous proteins (e.g. lipase, cellulase, carbohydrase) in	DR	WPI; 2004-707924/69.
CC	eukaryotic cells useful in industry to increase protein yields and to	DR	N-PSDB; ADS12800.
CC	facilitate purification. The present sequence is Trichoderma reesei HAC1	XX	
CC	protein. (Updated on 29-AUG-2003 to standardise OS field)	PT	Increasing the secretion of a heterologous protein, such as a therapeutic
XX		PT	or an industrial enzyme, in genetically modified eukaryotic cells by
SQ	Sequence 451 AA;	PT	inducing an elevated unfolded protein response (UPR).
Query Match		XX	Example 3; SEQ ID NO 2; 83bp; English.
Best Local Similarity		XX	The invention describes a method of increasing the secretion of a
Matches 451; Conservative 0; Mismatches 0; Indels 0; Gaps 0;		CC	heterologous protein in a eukaryotic cell, comprising inducing an
QY	1 MAFQSSPLVKFEASPAESFLSAPGDNFTSLPADSTPTLNPRDMTPDPSVADIDSRLSV 60	CC	elevated unfolded protein response (UPR). Also described are: an isolated
Db		CC	nucleic acid encoding a HAC1 protein, where the HAC1 protein induces
	1 MAFQSSPLVKFEASPAESFLSAPGDNFTSLPADSTPTLNPRDMTPDPSVADIDSRLSV 60	CC	unfolded protein response and has less than 50% similarity to yeast HAC1
QY	61 IPESQDAEDDESHSATSATPSTSEKKPVKKRSGQVLPEPKTNLPKRKRAKTEDEKEOR 120	CC	protein; an isolated nucleic acid encoding a HAC1 protein, where the HAC1
Db		CC	protein induces unfolded protein response and where the HAC1 protein
	61 IPESQDAEDDESHSATSATPSTSEKKPVKKRSGQVLPEPKTNLPKRKRAKTEDEKEOR 120	CC	comprises a DNA binding region that has greater than 70% similarity to
QY	121 RVERVLNRRAAQSSRRERKRLEVEALEKRNKELETLLINVQKTNLILVEELNFRFRSSGV 180	CC	the DNA binding region of filamentous fungi HAC1 protein; a protein
Db		CC	having unfolded protein response inducing activity and having greater
	121 RVERVLNRRAAQSSRRERKRLEVEALEKRNKELETLLINVQKTNLILVEELNFRFRSSGV 180	CC	than 70% similarity to a fully defined amino acid sequence of 451, 349 or
QY	241 PISDKFEQTKEDBEQADEMEQTHETKAAAKKNSQSRVSTDTORPAVSIGG 300	CC	386, or 504 or 1232 amino acids (SEQ ID NO: 2, 4 or 16, or 12 or 14,
Db		CC	respectively) as given in the specification; a protein having an amino
	241 PISDKFEQTKEDBEQADEMEQTHETKAAAKKNSQSRVSTDTORPAVSIGG 300	CC	acid sequence of SEQ ID NO: 2, 4, 8, 10, 12, 14 or 16; an isolated
QY	301 DAAVPFVSDDAGANCLGLDPVHQDDGPPFSIGHSGFLSAALDADRYLLESQLLASPNASTV 240	CC	nucleic acid encoding a PTC2 protein that modulates unfolded protein
Db		CC	response and has at least 70% similarity to a fully defined amino acid
	301 DAAVPFVSDDAGANCLGLDPVHQDDGPPFSIGHSGFLSAALDADRYLLESQLLASPNASTV 240	CC	sequence of 130 or 438 amino acids (SEQ ID NO: 8 or 10) as given in the
QY	361 DDYLAGDSAACTNPLPSDYDFDINDFLTDDANHAAYDIVAASNYYAADRELDLEIHDP 420	CC	specification; an isolated nucleic acid encoding a IRE1 protein that
Db		CC	modulates unfolded protein response and has at least 60% similarity to a
	361 DDYLAGDSAACTNPLPSDYDFDINDFLTDDANHAAYDIVAASNYYAADRELDLEIHDP 420	CC	fully defined amino acid sequence with SEQ ID NO: 12 or 14; a protein
QY	421 ENQIPSRHSIQPQSGASSHGCGDGGIAGV 451	CC	having unfolded protein response modulating activity and having greater
Db		CC	than 70% similarity to SEQ ID NO: 8 or 10; and a cell containing a
	421 ENQIPSRHSIQPQSGASSHGCGDGGIAGV 451	CC	heterologous nucleic acid encoding a protein having unfolded protein
RESULT 2		CC	response modulating activity and a heterologous nucleic acid encoding a
ADS12801		CC	protein of interest to be secreted. The methods and compositions of
ID	ADS12801 standard; protein; 451 AA.	CC	genetically manipulating cells to have an elevated unfolded protein
XX	AC	CC	response (UPR) resulting in an increased capacity to produce secreted
XX	AC	CC	proteins, are useful in e.g. production of therapeutic or industrial
XX	16-DEC-2004 (first entry)	CC	enzymes. This is the amino acid sequence of Trichoderma reesei HAC1
XX	Trichoderma reesei hac1 chaperone and foldase.	XX	chaperone and foldase.
DE		SQ	Sequence 451 AA;
XX		Query Match	
KW	unfolded protein response; UPR; HAC1; PTC2;	Best Local Similarity	
KW	unfolded protein response modulator; enzyme production; hac1; chaperone;	Matches 451; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
KW	foldase; enzyme.		
XX		QY	
OS	Hypocrea jecorina.	Db	
XX		QY	
PN	US2004186070-A1.	Db	
XX		QY	
PD	23-SEP-2004.	Db	
XX		QY	
PF	15-SEP-2003; 2003US-00663450.	Db	
XX		QY	
PR	24-MAR-2000; 2000US-00534692.	Db	
PR	23-MAR-2001; 2001US-00816277.	QY	
XX		Db	
PA	(GENV) GENENCOR INT INC.	QY	
XX		Db	
PI	Penttila ME, Ward M, Wang H, Valkonen MJ, Saloheimo MJA;	QY	
XX		Db	

QY 421 ENQIPSRHSIQPOSGASSHGCDGGIAGV 451
 Db 421 ENQIPSRHSIQPOSGASSHGCDGGIAGV 451

RESULT 3

ADSL2804
 ID ADSL2804 standard; protein; 451 AA.

XX AC ADSL2804;

XX DT 16-DEC-2004 (first entry)

XX DE Trichoderma reesei hac1 DNA binding domain.

XX KW unfolded protein response; UPR; HAC1; PTC2;

XX KW unfolded protein response modulator; enzyme production; hac1; chapertone;

XX KW foldase; DNA binding domain.

XX OS Hypocrea jecorina.

XX PN US2004186070-A1.

XX PD 23-SEP-2004.

XX PF 15-SEP-2003; 2003US-00663450.

XX PR 24-MAR-2000; 2000US-00534692.

XX PR 23-MAR-2001; 2001US-00816277.

XX PA (GENV) GENENCOR INT INC.

XX PI Penttila ME, Ward M, Wang H, Valkonen MJ, Saloheimo MLA;

XX DR WPI; 2004-707924/69.

XX PT Increasing the secretion of a heterologous protein, such as a therapeutic
 PT or an industrial enzyme, in genetically modified eukaryotic cells by
 PT inducing an elevated unfolded protein response (UPR).

XX PS Example 3; SEQ ID NO 5; 83pp; English.

XX CC The invention describes a method of increasing the secretion of a
 CC heterologous protein in a eukaryotic cell, comprising inducing an
 CC elevated unfolded protein response (UPR). Also described are: an isolated
 CC nucleic acid encoding a HAC1 protein, where the HAC1 protein induces
 CC unfolded protein response and has less than 50% similarity to yeast HAC1
 CC protein; an isolated nucleic acid encoding a HAC1 protein, where the HAC1
 CC protein induces unfolded protein response and where the HAC1 protein
 CC comprises a DNA binding region that has greater than 70% similarity to
 CC the DNA binding region of filamentous fungi HAC1 protein; a protein
 CC having unfolded protein response inducing activity and having greater
 CC than 70% similarity to a fully defined amino acid sequence of 451, 349 or
 CC 386, or 504 or 1232 amino acids (SEQ ID NO: 2, 4 or 16, or 12 or 14,
 CC respectively) as given in the specification; a protein having an amino
 CC acid sequence of SEQ ID NO: 2, 4, 8, 10, 12, 14 or 16; an isolated
 CC nucleic acid encoding a PTC2 protein that modulates unfolded protein
 CC response and has at least 70% similarity to a fully defined amino acid
 CC sequence of 130 or 438 amino acids (SEQ ID NO: 8 or 10) as given in the
 CC specification; an isolated nucleic acid encoding a IRE1 protein that
 CC modulates unfolded protein response and has at least 60% similarity to a
 CC fully defined amino acid sequence with SEQ ID NO: 12 or 14; a protein
 CC having unfolded protein response modulating activity and having greater
 CC than 70% similarity to SEQ ID NO: 8 or 10; and a cell containing a
 CC heterologous nucleic acid encoding a protein having unfolded protein
 CC response modulating activity and a heterologous nucleic acid encoding a
 CC protein of interest to be secreted. The methods and compositions of
 CC genetically manipulating cells to have an elevated unfolded protein
 CC response (UPR) resulting in an increased capacity to produce secreted
 CC proteins, are useful in e.g. production of therapeutic or industrial
 CC enzymes. This is the amino acid sequence of Trichoderma reesei HAC1
 CC chapertone and foldase DNA binding domain.

XX SQ Sequence 451 AA;

Query Match 100.0%; Score 2294; DB 8; Length 451;
 Best Local Similarity 100.0%; Pred. No. 3.7e-179;
 Matches 451; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MAFQSSPLVKFEASPAESFLSAPGDNFTSLPADSTLNPDRMTTPDSVADIDSRLSV 60
 Db 1 MAFQSSPLVKFEASPAESFLSAPGDNFTSLPADSTLNPDRMTTPDSVADIDSRLSV 60
 QY 61 IPESQDAEDDESHSTSATAPSTSEKKPVKKRKSQGVLPKPTNLNPPKRAKTEDEKEOR 120
 Db 61 IPESQDAEDDESHSTSATAPSTSEKKPVKKRKSQGVLPKPTNLNPPKRAKTEDEKEOR 120
 QY 121 RVERVLNRRAAQSSRRERKLEVEALEKRNKELETLLINVQKTNLLI LVEELNFRSSGV 180
 Db 121 RVERVLNRRAAQSSRRERKLEVEALEKRNKELETLLINVQKTNLLI LVEELNFRSSGV 180
 QY 181 VTRSSSPLDSLODSITLSQOLFCSRDCGQTMSPQSLMDQIMRSAANPTVNPASLSPSLP 240
 Db 181 VTRSSSPLDSLODSITLSQOLFCSRDCGQTMSPQSLMDQIMRSAANPTVNPASLSPSLP 240
 QY 241 PISDKFEQTKKEDEDEQADEMEQTHETKEAAAAKEKNSKOSRVSTDTORPAVSIGG 300
 Db 241 PISDKFEQTKKEDEDEQADEMEQTHETKEAAAAKEKNSKOSRVSTDTORPAVSIGG 300
 QY 301 DAAVPVFSDDAGANCLGLDPVHODDGPFSIGHSGLSAALDADRYLLESOLLASPNASTV 360
 Db 301 DAAVPVFSDDAGANCLGLDPVHODDGPFSIGHSGLSAALDADRYLLESOLLASPNASTV 360
 QY 361 DDYLAGDSAACTNPDPDYDFDINDFLTDDANHAAYDIVAASNYAAANR...THDP 420
 Db 361 DDYLAGDSAACTNPDPDYDFDINDFLTDDANHAAYDIVAASNYAAANR...THDP 420
 QY 421 ENQIPSRHSIQPOSGASSHGCDGGIAGV 451
 Db 421 ENQIPSRHSIQPOSGASSHGCDGGIAGV 451

RESULT 4

AAB82975

ID AAB82975 standard; protein; 450 AA.

XX AC AAB82975;

XX DT 11-SEP-2003 (revised)

DT 21-DEC-2001 (first entry)

XX OS Trichoderma reesei HAC1, involved in unfolded protein response

XX KW HAC1; transcription factor; unfolded protein response; protein secretion.

XX OS Hypocrea jecorina.

XX FT Key Location/Qualifiers
 FT Domain 84..147
 FT /label= DNA binding domain

XX PN WO200172783-A2.

XX PD 04-OCT-2001.

XX PF 23-MAR-2001; 2001WO-US009401.

XX PR 24-MAR-2000; 2000US-00534692.

XX PA (GENV) GENENCOR INT INC.

XX PI Penttila ME, Ward M, Wang H, Valkonen MJ, Saloheimo MLA;

XX DR WPI; 2001-626252/72.

XX DR N-PSDB; AAB26931.

XX Increasing secretion of heterologous proteins e.g. lipase and cellulase
PT in eukaryotic cells useful in industry to increase production and
PT facilitate purification, by inducing an elevated unfolded protein
PT response.
XX
PS Claim 54; Fig 7A-B; 89pp; English.
XX
CC The present sequence is that of the HAC1 protein of *Trichoderma reesei*,
CC as deduced from the newly isolated HAC1 gene (see AAH26931). HAC1 protein
CC is a transcription factor involved in the unfolded protein response
CC (UPR). The invention provides methods for increasing the secretion of a
CC heterologous protein in a cell by inducing an elevated UPR. This can be
CC achieved by modulating the activity of HAC1, PTC2 or IRE1 in the cell,
CC e.g. by gene overexpression. The cell from which the protein is secreted
CC can be any cell having an UPR, such as mammalian cells, insect cells,
CC yeast and filamentous fungi. The protein of interest can be any secreted
CC protein such as a therapeutic protein or an industrial enzyme, e.g.
CC lipase, cellulase, endoglucanase-H, protease, carbohydratase, reductase,
CC oxidase, isomerase, transferase, kinase, phosphatase, alpha-amylase,
CC glucosylase, lignocellulose hemicellulase, pectinase and ligninase
CC (claimed). (Updated on 11-SEP-2003 to standardise OS field)
XX
SQ Sequence 450 AA;

Query Match 99.4%; Score 2279.5; DB 4; Length 450;
Best Local Similarity 99.8%; Pred. No. 5.7e-178;
Matches 450; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 MAFQSSPLVKFEASPAESFLGAPGNFTSLFADSTPSTLNPRDMTPDSVADIDSRLSV 60
DB 1 MAFQSSPLVKFEASPAESFLGAPGNFTSLFADSTPSTLNPRDMTPDSVADIDSRLSV 60
QY 61 IPESQDAEDESHSTSATPSTSEKKPVKKRSGQVLPPEKTNLPKRKRAKTEDEKEOR 120
DB 61 IPESQDAEDESHSTSATPSTSEKKPVKKRSGQVLPPEKTNLPKRKRAKTEDEKEOR 120
QY 121 RVERVLNRRAAQSSRRERKLEVEALEKRNKELETLLINVQKTNLLVLELNRFRSSGV 180
DB 121 RVERVLNRRAAQSSRRERKLEVEALEKRNKELETLLINVQKTNLLVLELNRFRSSGV 180
QY 181 VTRSSSPLSLQDSITLSQOLFSGRDGQTMNPEQSLMDQIMRSAANPTVNPASLPSLP 240
DB 181 VTRSSSPLSLQDSITLSQOLFSGRDGQTMNPEQSLMDQIMRSAANPTVNPASLPSLP 240
QY 241 PISDKFEQTKEDDEQADDEMEQTHETKAAAKKRNKQSRVSTDTORPAVSGIG 300
DB 241 PISDKFEQTKEDDEQADDEMEQTHETKAAAKKRNKQSRVSTDTORPA-SIGG 299
QY 301 DAAVPVFSDDAGANCLGLDPVHQDDGPFSGHSGFLSAAALDADRYLLESQLLASPNASTV 360
DB 300 DAAVPVFSDDAGANCLGLDPVHQDDGPFSGHSGFLSAAALDADRYLLESQLLASPNASTV 359
QY 361 DDDYLAGDSAACTNPLPSDYDFDINDFLTDDANHAAYDIVAASNYAAADRELIDLHDP 420
DB 360 DDDYLAGDSAACTNPLPSDYDFDINDFLTDDANHAAYDIVAASNYAAADRELIDLHDP 419
QY 421 ENQIPRHSITQOPQSGASSHGCGDGGIAGV 451
DB 420 ENQIPRHSITQOPQSGASSHGCGDGGIAGV 450

RESULT 5
AAB82976
ID AAB82976 standard; protein; 350 AA.
XX
AC AAB82976;
XX
DT 11-SEP-2003 (revised)
DT 21-DEC-2001 (first entry)
XX
DE *Aspergillus nidulans* hacA, involved in unfolded protein response.
XX

KW HAC1; transcription factor; unfolded protein response; protein secretion.
XX
OS *Emmericella nidulans*.
XX
FH Key Location/Qualifiers
FT Domain 53..116
FT /label= DNA binding domain
XX
PN WO200172783-A2.
XX
PD 04-OCT-2001.
XX
PF 23-MAR-2001; 2001WO-US009401.
XX
PR 24-MAR-2000; 2000US-00534692.
XX
PA (GEMV) GENENCOR INT INC.
XX
PI Penttila ME, Ward M, Wang H, Valkonen MJ, Saikholov M A,
XX
XX WPI; 2001-626252/72.
DR N-PSDB; AAH26932.
XX
XX Increasing secretion of heterologous proteins e.g. lipase and cellulase
PT in eukaryotic cells useful in industry to increase production and
PT facilitate purification, by inducing an elevated unfolded protein
PT response.
XX
PS Claim 54; Fig 8A-B; 89pp; English.
XX
CC The present sequence is that of the hacA protein of *Aspergillus nidulans*,
CC as deduced from the newly isolated hacA gene (see AAH26932). HacA protein
CC is a transcription factor involved in the unfolded protein response
CC (UPR). The invention provides methods for increasing the secretion of a
CC heterologous protein in a cell by inducing an elevated UPR. This can be
CC achieved by modulating the activity of HAC1 (or hacA), PTC2 or IRE1 in
CC the cell, e.g. by gene overexpression. The cell from which the protein is
CC secreted can be any cell having an UPR, such as mammalian cells, insect
CC cells, yeast and filamentous fungi. The protein of interest can be any
CC secreted protein such as a therapeutic protein or an industrial enzyme,
CC e.g. lipase, cellulase, endoglucanase-H, protease, carbohydratase,
CC reductase, oxidase, isomerase, transferase, kinase, phosphatase, alpha-
CC amylase, glucosylase, lignocellulose hemicellulase, pectinase and
CC ligninase (claimed). (Updated on 11-SEP-2003 to standardise OS field)
XX
SQ Sequence 350 AA;

Query Match 19.1%; Score 437.5; DB 4; Length 350;
Best Local Similarity 32.4%; Pred. No. 4e-27;
Matches 146; Conservative 51; Mismatches 122; Indels 131; Gaps 18;

QY 26 DNFTSL----FADSTPSTLNPRDMTPDSVADIDSRLSVIPESQDAEDESHSTSATAP 80
DB 5 DRFPVKMEDAFANSLPTTPPSLEVPLTVSPADTSLRTKNV-----VAOTKP 51
QY 81 STSEKKPVKKRSGQVLPPEKTNLPKRKRAKTEDEKEQRRVERVLNRRAAQSSRRERK 140
DB 52 --EKKPAKKRSGQVLPPEKTNLPKRKRAKTEDEKEQRRVERVLNRRAAQSSRRERK 104
QY 141 LEVEALEKRNKELETLLINVQKTNLLVLELNRFRSSGVVTRSSSPIDSHJDSITTCQ 200
DB 110 LEMEKLESEK-----IDMEQNOQFLQRLQAQMAENNRLS-----OOVAQLSAF 141
QY 201 LFGRSDGQTMNPEQSLMDQIMRSAANPTVNPASLPSLPISDKFEQTKEDDEQADDE 459
DB 154 VRGSRH----STPTSS-----SPASVSTLPTLTKFO----EGDEVPLDRI 191
QY 260 --DEMEQTHETKAAAKKRNKQSRVSTDTORPAVSGIG-----DAAVPVFSDDAGA 313
DB 192 PPTPSVTSDYSPTLKPSSLAEE-----SPDLTQHPAVSVGGLEGESALTFLF--DLGA 241
QY 314 NCLGLDPVHQDDGPFSS---IGHSGFLSAAALDADRYLLE-----SOLLASPNASTV 360
DB 314 NCLGLDPVHQDDGPFSS---IGHSGFLSAAALDADRYLLE-----SOLLASPNASTV 360

Db 242 S-IKHEPTHDLTAPLSDDDFRLLFNGDSLSLESDSSLEDDGPAFDVLDGDLGSAFPFDSMV 300
 QY 361 DDYLAGDSAACTNPLPSDYDFDINDFLTDANHAAYDIVAASNYAAADRELDLEIHPD 420
 Db 301 DFD-----TEPVTLLE-DLEQTNGLSDSASCKAASL----- 329
 QY 421 ENQIPSRHSIQPOSGASHGCDGGIAGV 450
 Db 330 -----QPSHGASTSRCDGQGIAG 348

RESULT 6
 AAE15372
 ID AAE15372 standard; protein; 350 AA.
 XX AC AAE15372;
 XX 29-AUG-2003 (revised)
 DT 07-MAR-2002 (first entry)
 XX
 DE Aspergillus nidulans hacA protein.
 XX
 KW Heterologous protein secretion; unfolded protein response; UPR; lipase;
 KW cellulase; carbohydrase; industry; purification; hacA protein.
 OS Emericella nidulans.
 XX
 FH Key Location/Qualifiers
 FT Binding-site 53..116
 FT /label= DNA-binding_domain
 XX
 PN US2001034045-A1.
 XX
 PD 25-OCT-2001.
 XX
 PF 23-MAR-2001; 2001US-00816277.
 XX
 PR 24-MAR-2000; 2000US-00534692.
 XX
 PA (GENM) GENENCOR INT INC.
 XX
 PI Penttila ME, Ward M, Wang H, Valkonen MJ, Saloheimo MLA;
 XX WPI; 2002-033728/04.
 DR N-PSDB; AAD24596.
 XX
 PT Increasing secretion of heterologous proteins e.g. lipase and cellulase
 PT in eukaryotic cells useful in industry to increase production and
 PT facilitate purification, by inducing an elevated unfolded protein
 PT response.
 XX
 PS Claim 39; Fig 8; 56pp; English.
 XX
 CC The present invention relates to methods for increasing the secretion of
 CC heterologous protein in eukaryotic cells by inducing an elevated unfolded
 CC protein response (UPR). The method involves inducing the elevated UPR by
 CC increasing the presence of proteins such as HAC1, HACA, PTC2 or IRE1 in
 CC cells. The method and sequences are useful for increasing the secretion
 CC of heterologous proteins (e.g. lipase, cellulase, carbohydrase) in
 CC eukaryotic cells useful in industry to increase protein yields and to
 CC facilitate purification. The present sequence is Aspergillus nidulans
 CC hacA protein. (Updated on 29-AUG-2003 to standardise OS field)
 XX
 SQ Sequence 350 AA;
 Query Match 19.1%; Score 437.5; DB 5; Length 350;
 Best Local Similarity 32.4%; Pred. No. 48-27;
 Matches 146; Conservative 51; Mismatches 122; Indels 131; Gaps 18;
 QY 26 DNFTSL-----FADSTPTLNPRDMTPDSVADIDSRLSVIPESQDAEDSHSATSAP 80
 Db 5 DRFSPVKMEDAFANSLPTTPSLVFLVTVSPADTSILRTKNV-----VAQTKP 51

QY 81 STSEKKPVKKRSGWGLPEPKTNLPKRKRAKTEDEKQRRVERVLNRRAAQSRRERK 140
 Db 52 --EEKKPAKKRSGWGLPEPKTNLPKRKRAKTEDEKQRRVERVLNRRAAQSRRERK 109
 QY 141 LEVEALEKRNKELETLINVOKTNLILVEELNRRFRSSGVVTRSSPLDSLQDSITLSQO 200
 Db 110 LEMEKLESEK-----IDMEQONQLLQRLQAEMENRLS-----QQVAQLSAE 153
 QY 201 LFGSRDQGTMSNPESLMDQIMRSAAANTVTNPASLSPSPPIPSDKFQTKEDDEQADE- 259
 Db 154 VRGSRH-----STPTSS-----SPASVSPTLTPTLFKQ-----EGDEVPLDRI 191
 QY 260 --DEMEQOTWHTKEAAAAAKEKNQSRVSTDTORPAVSIIGG-----DAAPVFESDACA 313
 Db 192 PPTPSVTDYSPFLKPPSLAE-----SPDLTQHPAVSVGGLGDESDALTLF--DLGA 241
 QY 314 NCLGLDPVHQDDGPFS----IGHFGLSAAALDADRYLLE-----SOLLASPNASTV 360
 Db 242 S-IKHEPTHDLTAPLSDDDFRLLFNGDSLSLESDSSLEDDGPAFDVLDGDLGSAFPFDSMV 300
 QY 361 DDYLAGDSAACTNPLPSDYDFDINDFLTDANHAAYDIVAASNYAAADRELDLEIHPD 420
 Db 301 DFD-----TEPVTLLE-DLEQTNGLSDSASCKAASL----- 329
 QY 421 ENQIPSRHSIQPOSGASHGCDGGIAGV 450
 Db 330 -----QPSHGASTSRCDGQGIAG 348

RESULT 7
 AAE15379
 ID AAE15379 standard; protein; 386 AA.
 XX AC AAE15379;
 XX 07-MAR-2002 (first entry)
 DT
 DE Aspergillus niger var. awamori hacA protein #1.
 XX
 KW Heterologous protein secretion; unfolded protein response; UPR; lipase;
 KW cellulase; carbohydrase; industry; purification; hacA protein.
 OS Aspergillus niger.
 XX
 PN US2001034045-A1.
 XX
 PD 25-OCT-2001.
 XX
 PF 23-MAR-2001; 2001US-00816277.
 XX
 PR 24-MAR-2000; 2000US-00534692.
 XX
 PA (GENM) GENENCOR INT INC.
 XX
 PI Penttila ME, Ward M, Wang H, Valkonen MJ, Saloheimo MLA;
 XX WPI; 2002-033728/04.
 DR N-PSDB; AAD24601.
 XX
 PT Increasing secretion of heterologous proteins e.g. lipase and cellulase
 PT in eukaryotic cells useful in industry to increase production and
 PT facilitate purification, by inducing an elevated unfolded protein
 PT response.
 XX
 PS Claim 39; Fig 28; 56pp; English.
 XX
 CC The present invention relates to methods for increasing the secretion of
 CC heterologous protein in eukaryotic cells by inducing an elevated unfolded
 CC protein response (UPR). The method involves inducing the elevated UPR by
 CC increasing the presence of proteins such as HAC1, HACA, PTC2 or IRE1 in
 CC cells. The method and sequences are useful for increasing the secretion
 CC of heterologous proteins (e.g. lipase, cellulase, carbohydrase) in
 CC eukaryotic cells useful in industry to increase protein yields and to

```
CC facilitate purification. The present sequence is Aspergillus niger var.
CC awamori hacA protein
XX
SQ Sequence 386 AA;

Query Match      18.9%; Score 434.5; DB 5; Length 386;
Best Local Similarity 32.3%; Pred. No. 9e-27;
Matches 147; Conservative 50; Mismatches 139; Indels 119; Gaps 20;

QY 24 PGDNFTSLFADSTP-STLNPRDMTPDSDVADIDS-----RLSVIPESQDAEDDES 72
Db 22 PGTTPGLAPPASPVMTWVPVFLMMEAFSPVDSLGSAGPTPELPPLTVSPADTSLDSDS-- 79
QY 73 HSTSATAPSTSEKKPKVKRKGQVLPEPKTNLPKRKAkteDEKEQRRVERVLNRRAA 132
Db 80 -SVQAGETKABEKKPKVKRKGQELPVKTNLPKRKAkteDEKEQRIERVLNRRAA 138
QY 133 QSSRRKRLEVEALEKRKNELETLLINVKTNLLIIVEELNFRSSGGVTVSSSPDLSLQ 192
Db 139 QTSRRKRLEMEKLENEK-----IQMEQNOQFLRLSQMEANNRLN-----Q 182
QY 193 DSITLSQQLFGSRDQOTMSNPESQSLMDQIMRSANPTVNPASLSPSLPPIIDKKEFQTEE 252
Db 183 QVAQLSAEVRGSR-GNT---PKPG-----SPVSASPTLTPTLFKQ-----ER 220
QY 253 DB---EQADEDEMEQTHWETKEAAAAAKEKNSKQSRVSTDTQRPAVSIIGG-----DAAVP 305
Db 221 DEIPLERIPFPPTPSITDYSPTLRPSTLAE-----SSDVTHQPAVAVAGLEGSGALS 272
QY 306 VPSDDAGANCLGLDPVHODDGPFGSHSGFGLSAAALDAURY--LLESQLLASPNASTVDD- 362
Db 273 LF--DVGSN---PEPHAADD-----LAAPLSDDDDFHLRFNVDSFVSGSDSSVLEDG 317
QY 363 ---DYLAGDSAACTNPLPSDYDFDINDFLTDDANHAAYDIVAASNYAAADRELDEIHD 419
Db 318 PAFDVLDDGGDLSAF--PFDSMVDFD-----PESVGFEGIE 350
QY 420 PENQIPSRHSIQ-----QPOSGASSHCDDGGGIAGV 450
Db 351 PPHGLPDETSRTSSVQPSLGASTSCDGGGIAG 385

RESULT 8
ID ADS12815 standard; protein; 386 AA.
XX ADS12815;
AC ADS12815;
XX
XX 16-DEC-2004 (first entry)
XX
XX Aspergillus nidulans hacA chaperone and foldase #3.
DE
XX unfolded protein response; UPR; HAC1; PTC2;
KW unfolded protein response modulator; enzyme production; hacA; chaperone;
KW foldase; enzyme.
XX
XX Emericella nidulans.
OS
XX
XX US2004186070-A1.
PN
XX
XX 23-SEP-2004.
XX
XX 15-SEP-2003; 2003US-00663450.
XX
XX 24-MAR-2000; 2000US-00534692.
PR
XX 23-MAR-2001; 2001US-00816277.
XX
XX (GEMV ) GENENCOR INT INC.
PA
XX Penttila ME, Ward M, Wang H, Valkonen MJ, Saloheimo MIA;
PI WPI; 2004-707924/69.
XX
XX N-PSDB; ADS12814.
DR
```

XX Increasing the secretion of a heterologous protein, such as a thermostable or an industrial enzyme, in genetically modified eukaryotic cells by inducing an elevated unfolded protein response (UPR).

XX Example 12; SEQ ID NO 16; 83pp; English.

XX The invention describes a method of increasing the secretion of a heterologous protein in a eukaryotic cell, comprising inducing an elevated unfolded protein response (UPR). Also described are: an isolated nucleic acid encoding a HAC1 protein, where the HAC1 protein induces unfolded protein response and has less than 50% similarity to yeast HAC1 protein; an isolated nucleic acid encoding a HAC1 protein, where the HAC1 protein induces unfolded protein response and where the HAC1 protein comprises a DNA binding region that has greater than 70% similarity to the DNA binding region of filamentous fungi HAC1 protein; a protein having unfolded protein response inducing activity and having at least 70% similarity to a fully defined amino acid sequence of SEQ ID NO: 2, 4, 8, 10, 12, 14 or 16; an isolated nucleic acid sequence of SEQ ID NO: 2, 4 or 16, or 12 or 14, 386, or 504 or 1232 amino acids (SEQ ID NO: 2, 4 or 16, or 12 or 14, respectively) as given in the specification; a protein having an amino acid sequence of SEQ ID NO: 2, 4, 8, 10, 12, 14 or 16; an isolated nucleic acid encoding a PTC2 protein that modulates unfolded protein response and has at least 70% similarity to a fully defined amino acid sequence of 130 or 438 amino acids (SEQ ID NO: 8 or 10, respectively); a specification; an isolated nucleic acid encoding a HAC1 protein that modulates unfolded protein response and has at least 70% similarity to a fully defined amino acid sequence with SEQ ID NO: 12 or 14; a protein having unfolded protein response modulating activity and having greater than 70% similarity to SEQ ID NO: 8 or 10; and a cell containing a heterologous nucleic acid encoding a protein having unfolded protein response modulating activity and a heterologous nucleic acid encoding a protein of interest to be secreted. The methods and compositions of genetically manipulating cells to have an elevated unfolded protein response (UPR) resulting in an increased capacity to produce secreted proteins, are useful in e.g. production of therapeutic and industrial enzymes. This is the amino acid sequence of Aspergillus nidulans hacA chaperone and foldase.

XX Sequence 386 AA;

Query Match 18.9%; Score 434.5; DB 8; Length 386;
Best Local Similarity 32.3%; Pred. No. 9e-27;
Matches 147; Conservative 50; Mismatches 139; Indels 119; Gaps 20;

QY 24 PGDNFTSLFADSTP-STLNPRDMTPDSDVADIDS-----RLSVIPESQDAEDDES 72
Db 22 PGTTPGLAPPASPVMTWVPVFLMMEAFSPVDSLGSAGPTPELPPLTVSPADTSLDSDS-- 79
QY 73 HSTSATAPSTSEKKPKVKRKGQVLPEPKTNLPKRKAkteDEKEQRRVERVLNRRAA 132
Db 80 -SVQAGETKABEKKPKVKRKGQELPVKTNLPKRKAkteDEKEQRIERVLNRRAA 138
QY 133 QSSRRKRLEVEALEKRKNELETLLINVKTNLLIIVEELNFRSSGGVTVSSSPDLSLQ 192
Db 139 QTSRRKRLEMEKLENEK-----IQMEQNOQFLRLSQMEANNRLN-----Q 182
QY 193 DSITLSQQLFGSRDQOTMSNPESQSLMDQIMRSANPTVNPASLSPSLPPIIDKKEFQTEE 252
Db 183 QVAQLSAEVRGSR-GNT---PKPG-----SPVSASPTLTPTLFKQ-----ER 220
QY 253 DB---EQADEDEMEQTHWETKEAAAAAKEKNSKQSRVSTDTQRPAVSIIGG-----DAAVP 305
Db 221 DEIPLERIPFPPTPSITDYSPTLRPSTLAE-----SSDVTHQPAVAVAGLEGSGALS 272
QY 306 VPSDDAGANCLGLDPVHODDGPFGSHSGFGLSAAALDAURY--LLESQLLASPNASTVDD- 362
Db 273 LF--DVGSN---PEPHAADD-----LAAPLSDDDDFHLRFNVDSFVSGSDSSVLEDG 317
QY 363 ---DYLAGDSAACTNPLPSDYDFDINDFLTDDANHAAYDIVAASNYAAADRELDEIHD 419
Db 318 PAFDVLDDGGDLSAF--PFDSMVDFD-----PESVGFEGIE 350
QY 420 PENQIPSRHSIQ-----QPOSGASSHCDDGGGIAGV 450

Db	351	PPHGLPDETSRQTSSVQPSLGASTRCDGGIAAG	385
RESULT 9			
ADSL2805			
ID	ADSL2805	standard; protein; 349 AA.	
XX	AC	ADSL2805;	
XX	DT	16-DEC-2004 (first entry)	
XX	DE	Aspergillus nidulans haca DNA binding domain.	
XX	KW	unfolded protein response; UPR; HAC1; PTC2;	
XX	KW	unfolded protein response modulator; enzyme production; haca; chaperone;	
XX	KW	foldase; DNA binding domain.	
XX	OS	Emericella nidulans.	
XX	PN	US2004186070-A1.	
XX	PD	23-SEP-2004.	
XX	PF	15-SEP-2003; 2003US-00663450.	
XX	PR	24-MAR-2000; 2000US-00534692.	
XX	PR	23-MAR-2001; 2001US-00816277.	
XX	PA	(GEMV) GENENCOR INT INC.	
XX	PI	Penttila ME, Ward M, Wang H, Valkonen MJ, Saloheimo MLA;	
XX	DR	WPI; 2004-707924/69.	
XX	PT	Increasing the secretion of a heterologous protein, such as a therapeutic or an industrial enzyme, in genetically modified eukaryotic cells by inducing an elevated unfolded protein response (UPR).	
XX	PS	Example 3; SEQ ID NO 6; 83pp; English.	
XX	CC	The invention describes a method of increasing the secretion of a heterologous protein in a eukaryotic cell, comprising inducing an elevated unfolded protein response (UPR). Also described are: an isolated nucleic acid encoding a HAC1 protein, where the HAC1 protein induces unfolded protein response and has less than 50% similarity to yeast HAC1 protein; an isolated nucleic acid encoding a HAC1 protein, where the HAC1 protein induces protein; an isolated nucleic acid encoding a HAC1 protein, where the HAC1 protein induces protein induces unfolded protein response and where the HAC1 protein comprises a DNA binding region that has greater than 70% similarity to the DNA binding region of filamentous fungi HAC1 protein; a protein having unfolded protein response inducing activity and having greater than 70% similarity to a fully defined amino acid sequence of 451, 349 or 386, or 504 or 1232 amino acids (SEQ ID NO: 2, 4 or 16, or 12 or 14, respectively) as given in the specification; a protein having an amino acid sequence of SEQ ID NO: 2, 4, 8, 10, 12, 14 or 16; an isolated nucleic acid encoding a PTC2 protein that modulates unfolded protein response and has at least 70% similarity to a fully defined amino acid sequence of 130 or 438 amino acids (SEQ ID NO: 8 or 10) as given in the specification; an isolated nucleic acid encoding a IRE1 protein that modulates unfolded protein response and has at least 60% similarity to a fully defined amino acid sequence with SEQ ID NO: 12 or 14; a protein having unfolded protein response modulating activity and having greater than 70% similarity to SEQ ID NO: 8 or 10; and a cell containing a heterologous nucleic acid encoding a protein having unfolded protein response modulating activity and a heterologous nucleic acid encoding a protein of interest to be secreted. The methods and compositions of genetically manipulating cells to have an elevated unfolded protein response (UPR) resulting in an increased capacity to produce secreted proteins, are useful in e.g. production of therapeutic or industrial enzymes. This is the amino acid sequence of Aspergillus nidulans HAC1 chaperone and foldase DNA binding domain.	
XX	SQ	Sequence 349 AA;	

```
XX PS Example 3; SEQ ID NO 4; 83pp; English.
XX CC The invention describes a method of increasing the secretion of a
CC heterologous protein in a eukaryotic cell, comprising inducing an
CC elevated unfolded protein response (UPR). Also described are: an isolated
CC nucleic acid encoding a HAC1 protein, where the HAC1 protein induces
CC unfolded protein response and has less than 50% similarity to yeast HAC1
CC protein; an isolated nucleic acid encoding a HAC1 protein, where the HAC1
CC protein induces unfolded protein response and where the HAC1 protein
CC comprises a DNA binding region that has greater than 70% similarity to
CC the DNA binding region of filamentous fungi HAC1 protein; a protein
CC having unfolded protein response inducing activity and having greater
CC than 70% similarity to a fully defined amino acid sequence of 451, 349 or
CC 386, or 504 or 1232 amino acids (SEQ ID NO: 2, 4 or 16, or 12 or 14,
CC respectively) as given in the specification; a protein having an amino
CC acid sequence of SEQ ID NO: 2, 4, 8, 10, 12, 14 or 16; an isolated
CC nucleic acid encoding a PTC2 protein that modulates unfolded protein
CC response and has at least 70% similarity to a fully defined amino acid
CC sequence of 130 or 438 amino acids (SEQ ID NO: 8 or 10) as given in the
CC specification; an isolated nucleic acid encoding a IRE1 protein that
CC modulates unfolded protein response and has at least 60% similarity to a
CC fully defined amino acid sequence with SEQ ID NO: 12 or 14; a protein
CC having unfolded protein response modulating activity and having greater
CC than 70% similarity to SEQ ID NO: 8 or 10; and a cell containing a
CC heterologous nucleic acid encoding a protein having unfolded protein
CC response modulating activity and a heterologous nucleic acid encoding a
CC protein of interest to be secreted. The methods and compositions of
CC genetically manipulating cells to have an elevated unfolded protein
CC response (UPR) resulting in an increased capacity to produce secreted
CC proteins, are useful in e.g. production of therapeutic or industrial
CC enzymes. This is the amino acid sequence of Aspergillus nidulans HAC1
CC chaperone and foldase. Note: This sequence differs from ADS12863 given in
XX figure 8.
XX PS Sequence 349 AA;
XX CC
XX CC Query Match 18.9%; Score 434; DB 8; Length 349;
XX CC Best Local Similarity 32.4%; Pred. No. 7.6e-27;
XX CC Matches 149; Conservative 50; Mismatches 109; Indels 152; Gaps 21;
XX CC
QY 26 DNFTSL-----FADS--TPS-----TLNPRP--WMTDPSVADIISRLSVIPESQDAEDD 70
DB 5 DRFSVPMEDAFANSFTTPTSLFVFLVTSFADTSRTKXNVAAQ----- 47
QY 71 ESHSTSATAPSTSEKKPVKRSWGVLPPEPTNLPPRRKRAKTEDEKEQRRVERVLNR 130
DB 48 -----TKP--EEKPAKRRKSWGOELPVPTNLPPRKRAKTEDEKEQRRVERVLNR 98
QY 131 AAQSRERKRLVEALEKRNKELETLINQVKTNLILVEELNFRSSGVVTRSSPLDS 190
DB 99 AAQTSRERKRLWEKLESEK-----IDMEQQNQFLQRLAQMEANRLS----- 143
QY 191 LQDSITLSQQLSGRDGQTMSPQSLMDQIMRSAANPTVNPASLSPSLPPTSDKEFTQK 250
DB 144 -QOVAQLSAEVRGSRH-----STPTSS-----SPASVSTPTLPTLFQK---- 180
QY 251 EDEDEQADE--DEEMEQTWHETKGAATAAKENSKSQSRVSTDTQRPVAVSIGG----DAA 303
DB 181 EGDEVPDLRIPPTTSVTYDSTLKPSSLAE-----SPDLTQHPAVSVGLEGDESA 232
QY 304 VPVFSDDAGANCLGDPVHODGPFS---IGHSFGLSAALDADRYLLE-----SQ 350
DB 233 LTLF--DLGAS-IKHEPTHTDLTAPLSDDDFRRLFNGLSDSSLESDDLDEGFADFVLDSDG 289
QY 351 LLASPNASTVDDYLADGASACFTNPLPSDYDFDINDFTDDANHAAYDIVAASNYAAD 410
DB 290 LSAFPDMSVDFD-----TEPVTLE-DLEQTNGLSDSASCVAASL----- 328
QY 411 RELDLEIHPENQIPSRHSIQPQSGASSGHGCDGGIAYG 450
DB 329 -----QPSHGASTSRCDGGQIAAG 347
```

```
RESULT 11
ADS12863
ID ADS12863 standard; protein; 349 AA.
XX AC ADS12863;
XX DT 16-DEC-2004 (first entry)
XX DE Aspergillus nidulans hacA chaperone and foldase #2.
XX KW unfolded protein response; UPR; HAC1; PTC2;
XX KW unfolded protein response modulator; enzyme production; HACA; chaperone;
XX KW foldase; enzyme.
XX OS Emericella nidulans.
XX PN US2004186070-A1.
XX PD 23-SEP-2004.
XX PF 15-SEP-2003; 2003US-00663450.
XX PR 24-MAR-2000; 2000US-00534692.
XX PR 23-MAR-2001; 2001US-00816277.
XX PA (GENW ) GENENCOR INT INC.
XX PI Penttila ME, Ward M, Wang H, Valkonen MJ, Saloheimo MLA;
XX WPI; 2004-707924/69.
XX N-PSDB; ADS12802.
XX PT Increasing the secretion of a heterologous protein, such as a therapeutic
XX PT or an industrial enzyme, in genetically modified eukaryotic cells by
XX PT inducing an elevated unfolded protein response (UPR).
XX PS Example 3; Fig 8; 83pp; English.
XX CC The invention describes a method of increasing the secretion of a
XX CC heterologous protein in a eukaryotic cell, comprising inducing an
XX CC elevated unfolded protein response (UPR). Also described are: an isolated
XX CC nucleic acid encoding a HAC1 protein, where the HAC1 protein induces
XX CC unfolded protein response and has less than 50% similarity to yeast HAC1
XX CC protein; an isolated nucleic acid encoding a HAC1 protein, where the HAC1
XX CC protein induces unfolded protein response and where the HAC1 protein
XX CC comprises a DNA binding region that has greater than 70% similarity to
XX CC the DNA binding region of filamentous fungi HAC1 protein; a protein
XX CC having unfolded protein response inducing activity and having greater
XX CC than 70% similarity to a fully defined amino acid sequence of 451, 349 or
XX CC 386, or 504 or 1232 amino acids (SEQ ID NO: 2, 4 or 16, or 12 or 14,
XX CC respectively) as given in the specification; a protein having an amino
XX CC acid sequence of SEQ ID NO: 2, 4, 8, 10, 12, 14 or 16; an isolated
XX CC nucleic acid encoding a PTC2 protein that modulates unfolded protein
XX CC response and has at least 70% similarity to a fully defined amino acid
XX CC sequence of 130 or 438 amino acids (SEQ ID NO: 8 or 10) as given in the
XX CC specification; an isolated nucleic acid encoding a IRE1 protein that
XX CC modulates unfolded protein response and has at least 60% similarity to a
XX CC fully defined amino acid sequence with SEQ ID NO: 12 or 14; a protein
XX CC having unfolded protein response modulating activity and having greater
XX CC than 70% similarity to SEQ ID NO: 8 or 10; and a cell containing a
XX CC heterologous nucleic acid encoding a protein having unfolded protein
XX CC response modulating activity and a heterologous nucleic acid encoding a
XX CC protein of interest to be secreted. The methods and compositions of
XX CC genetically manipulating cells to have an elevated unfolded protein
XX CC response (UPR) resulting in an increased capacity to produce secreted
XX CC proteins, are useful in e.g. production of therapeutic or industrial
XX CC enzymes. This is the amino acid sequence of Aspergillus nidulans HAC1
XX CC chaperone and foldase. Note: This sequence differs from ADS12803 given in
XX CC the sequence listing.
XX PS Sequence 349 AA;
```

```
Query Match      18.7%; Score 434; DB 8; Length 349;
Best Local Similarity 32.4%; Pred. No. 7.6e-27;
Matches 149; Conservative 50; Mismatches 109; Indels 152; Gaps 21;

QY 26 DNFTSL-----FADS--TFS-----TLNPRD--MMTPDSVADIDSRLSVIPESQDAEDD 70
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 5 DRFPVKMEDAFANSTPTTSLEVPVLTVPADTSLATKNVVAQ-----47

QY 71 ESHSTSATAPSTSEKPKVKKRKGQVLPKPTNLPKRKAKTEDEKEQRRVERVLNR 130
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 48 -----TKP--BEKPKAKRKGQVLPKPTNLPKRKAKTEDEKEQRRVERVLNR 98

QY 131 AAQSSRRERKLEVEALEKRNKELETLINVKTNLILVEELNFRSSGWVTRSSPLDS 190
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 99 AAGTSRRERKLEMEKLESEK-----IDMEQQNQLLQRLAQMEANNRLS-----143

QY 191 LQDSITLSQOLFGRDQGTMSNPESQSLMDQIMRASAANPTVNPASLSPSPPIPSIDKEFQTK 250
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 144 -QOVAQLSAEVRGSRH-----STPTSS-----SPASVSTLTPTLFKO-----180

QY 251 EDEEQADE---DEEMEQTWHETKEAAAKERNKSKOSRVSTSTORPAVSIGG---DAA 303
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 181 EGDEVPLDRIPFTPTSPVTDYSPTLKPSLAE-----SPDLTQHPAVSVGGLGEDSA 232

QY 304 VPVFSDDAGANCLGLDPVHQDDGPFSS---IGHSPGLSALDADRYLLE-----SQ 350
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 233 LTLF--DLGAS--IKHEPTHTDLTAPLSDDDFRRLFNGLSDSLESDDGFAFDVLDSDG 289

QY 351 LLAASPNASTVDDYLAGDSAACTNPLPSDYDPDINDFLTDDANHAAYDIVAASNYAAAD 410
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 290 LSAFPDSDNVDFD-----TEPVTLLE-DLEQTNGLSDSASCKAASL-----328

QY 411 RELDLEIHPENQIPSRHSIQPQSGASHGCDGDIAGV 450
   |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 329 -----QPSHGASTRCDCQGIAAG 347

RESULT 12
ABO43144
ID ABO43144 standard; protein; 409 AA.
XX
XX ABO43144;
XX
DT 23-SEP-2003 (first entry)
DE A. thaliana disease tolerance transcription factor, G1034.
XX
KW Plant; transcription factor; disease resistance; transgenic;
KW plant breeding; pathogens resistance; pests; resistance.
XX
OS Arabidopsis thaliana.
XX
XX US2003046723-A1.
XX
XX 06-MAR-2003.
XX
XX 22-MAR-2000; 2000US-00533029.
XX
XX 22-MAR-2000; 2000US-00533029.
XX
PA (HEAR/) HEARD J.
PA (BROU/) BROUN P.
PA (RIEC/) RIECHMANN J L.
PA (KEDD/) KEDDIE J.
PA (PINE/) PINEDA O.
PA (ADAM/) ADAM L.
PA (SANA/) SANAH R.
PA (ZHAN/) ZHANG J.
PA (YUGG/) YU G.
PA (RATC/) RATCLIFFE O.
PA (PILG/) PILGRIM M.
PA (JIAN/) JIANG C.
PA (REUB/) REUBER L.
```


Search completed: November 23, 2005, 03:18:37
Job time : 179.139 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 23, 2005, 03:12:51 ; Search time 6.65899 Seconds
(without alignments)
844.265 Million cell updates/sec

Title: US-10-663-450-60

Perfect score: 342

Sequence: 1 KSTLPKRAKTKKEQRR.....CSLENNLNSVNLEKLADHE 68

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: /cgn2_6/ptodata/1/iaa/5 COMB.pap.*
- 2: /cgn2_6/ptodata/1/iaa/6 COMB.pap.*
- 3: /cgn2_6/ptodata/1/iaa/H COMB.pap.*
- 4: /cgn2_6/ptodata/1/iaa/PTUS COMB.pap.*
- 5: /cgn2_6/ptodata/1/iaa/RE COMB.pap.*
- 6: /cgn2_6/ptodata/1/iaa/backfiles1.pap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	214	62.6	248	2	US-09-248-796A-18840
2	178.5	52.2	409	2	US-09-533-029-104
3	115	33.6	23	2	US-09-831-642-72
4	110	32.2	143	2	US-09-840-211A-1930
5	88.5	25.9	327	1	US-07-684-965-6
6	88.5	25.9	327	2	US-09-215-098-5
7	88.5	25.9	327	2	US-09-252-658-6
8	88.5	25.9	327	2	US-09-949-016-11202
9	87.5	25.6	68	2	US-09-299-495F-7
10	87.5	25.6	69	1	US-08-630-011A-7
11	87.5	25.6	92	2	US-09-299-495F-9
12	87.5	25.6	93	1	US-08-690-011A-9
13	86.5	25.3	315	2	US-09-949-016-7832
14	86	25.1	102	2	US-09-640-211A-1046
15	85.5	25.0	501	2	US-09-949-016-11281
16	84.5	24.7	521	1	US-08-721-684C-2
17	84.5	24.7	521	1	US-09-005-970-2
18	84.5	24.7	521	2	US-09-407-715-2
19	84	24.6	208	2	US-09-949-016-9078
20	84	24.6	295	1	US-08-343-443B-107
21	83	24.3	371	2	US-09-148-545-259
22	83	24.3	371	2	US-09-621-011-259
23	82.5	24.1	360	1	US-08-319-866-2
24	82.5	24.1	360	2	US-08-809-917-2
25	82.5	24.1	360	2	US-09-419-371-2
26	80	23.4	214	2	US-09-640-211A-1088
27	79.5	23.2	54	1	US-08-319-866-4

28	79.5	23.2	54	2	US-08-809-917-4	Sequence 4, Appli
29	79.5	23.2	54	2	US-09-419-371-4	Sequence 4, Appli
30	78.5	23.0	54	1	US-08-319-866-5	Sequence 5, Appli
31	78.5	23.0	54	2	US-08-809-917-5	Sequence 5, Appli
32	78.5	23.0	54	2	US-09-419-371-5	Sequence 5, Appli
33	78.5	23.0	351	2	US-09-870-089B-2	Sequence 2, Appli
34	78.5	23.0	362	2	US-09-949-016-10923	Sequence 10923, A
35	78	22.8	61	2	US-09-416-050A-26	Sequence 26, Appl
36	78	22.8	61	2	US-09-664-800-26	Sequence 26, Appl
37	78	22.8	61	2	US-09-865-309-26	Sequence 26, Appl
38	78	22.8	61	2	US-09-661-569-26	Sequence 26, Appl
39	78	22.8	307	2	US-09-267-031-6	Sequence 6, Appli
40	77.5	22.7	54	1	US-08-319-866-6	Sequence 6, Appli
41	77.5	22.7	54	2	US-08-809-917-6	Sequence 6, Appli
42	77.5	22.7	54	2	US-09-419-371-6	Sequence 6, Appli
43	77.5	22.7	71	2	US-09-270-767-50533	Sequence 50533, A
44	77.5	22.7	71	2	US-09-270-767-50533	Sequence 50533, A
45	76.5	22.4	196	2	US-09-640-211A-881	Sequence 881, App

ALIGNMENTS

RESULT 1

US-09-248-796A-18840
; Sequence 18840, Application US/09248796A
; Patent No. 6747137

; GENERAL INFORMATION:

; APPLICANT: Keith Weinstock et al

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICANS

; FILE REFERENCE: 107196.132

; CURRENT APPLICATION NUMBER: US/09/248,796A

; CURRENT FILING DATE: 1999-02-12

; PRIOR APPLICATION NUMBER: US 60/074,725

; PRIOR FILING DATE: 1998-02-13

; PRIOR APPLICATION NUMBER: US 60/096,409

; PRIOR FILING DATE: 1998-08-13

; NUMBER OF SEQ ID NOS: 28208

; SEQ ID NO 18840

; LENGTH: 248

; TYPE: PRT

; ORGANISM: Candida albicans

; FEATURE:

; NAME/KEY: UNSURE

; LOCATION: (221)

; OTHER INFORMATION: Identity of amino acid sequences at the above locations are unknown

US-09-248-796A-18840

Query Match 62.6%; Score 214; DB 2; Length 248;
Best Local Similarity 73.3%; Pred. No. 5e-16;
Matches 44; Conservative 5; Mismatches 11; Indels 0; Gaps 0;

QY	1	KSTLPKRAKTKKEQRRIRILNRRRAAHSREKRLHLQYLRKCSLENNLSVN 60
DB	72	KSTLPKRAKTKKEQRRIRILNRRRAAHSREKRLHLQYLRKCSLENNLSVN 131

RESULT 2

US-09-533-029-104
; Sequence 104, Application US/095333029
; Patent No. 6684446

; GENERAL INFORMATION:

; APPLICANT: Heard, Jacqueline

; APPLICANT: Broun, Pierre

; APPLICANT: Riechmann, Jose-Luis

; APPLICANT: Keddle, James

; APPLICANT: Pineda, Omlara

; APPLICANT: Adam, Luc

; APPLICANT: Samaha, Raymond

; APPLICANT: Zhang, James

; APPLICANT: Yu, Guo-Liang

; APPLICANT: Ratcliffe, Oliver

US-09-252-658-6

Query Match 25.9%; Score 88.5; DB 2; Length 327;
Best Local Similarity 33.9%; Pred. No. 0.047;
Matches 20; Conservative 16; Mismatches 22; Indels 1; Gaps 1;

Qy 6 PRKRAKTEKEORRIERILNRRAAHQSKREKKRLHLOYLERKCSLLENLLNSVLEKL 64
Db 258 PALPTQPAEAPRKREVLNMRRAECRRKKYKVCLENRAVAVLEN-QNKTLEL 315

RESULT 8

US-09-949-016-11202

Sequence 11202, Application US/09949016
Patent No. 6812339

GENERAL INFORMATION:
APPLICANT: VENTER, J. Craig et al.
TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
FILE REFERENCE: CLO01307
CURRENT APPLICATION NUMBER: US/09/949,016
CURRENT FILING DATE: 2000-04-14
PRIOR APPLICATION NUMBER: 60/241,755
PRIOR FILING DATE: 2000-10-20
PRIOR APPLICATION NUMBER: 60/237,768
PRIOR FILING DATE: 2000-10-03
PRIOR APPLICATION NUMBER: 60/231,498
PRIOR FILING DATE: 2000-09-08
NUMBER OF SEQ ID NOS: 207012
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 11202
LENGTH: 327
TYPE: PRT
ORGANISM: Human

US-09-949-016-11202

Query Match 25.9%; Score 88.5; DB 2; Length 327;
Best Local Similarity 33.9%; Pred. No. 0.047;
Matches 20; Conservative 16; Mismatches 22; Indels 1; Gaps 1;

Qy 6 PRKRAKTEKEORRIERILNRRAAHQSKREKKRLHLOYLERKCSLLENLLNSVLEKL 64
Db 258 PALPTQPAEAPRKREVLNMRRAECRRKKYKVCLENRAVAVLEN-QNKTLEL 315

RESULT 9

US-09-299-495F-7

Sequence 7, Application US/09299495F
Patent No. 6361968

GENERAL INFORMATION:
APPLICANT: VINSON, Charles R.
KRYLOV, Dmitry
TITLE OF INVENTION: EXTENSION OF A PROTEIN-PROTEIN
INTERACTION SURFACE TO INACTIVATE THE FUNCTION OF A
CELLULAR PROTEIN

NUMBER OF SEQUENCES: 64
CORRESPONDENCE ADDRESS:
ADDRESSER: MORGAN & FINNEGAN, L.L.P.
STREET: 345 Park Avenue
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10154-0053
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: MS WORD 97
SOFTWARE: ASCII

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/299,495F
FILING DATE: 26-Apr-1999
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/001,654

US-09-252-658-6

Query Match 25.9%; Score 88.5; DB 2; Length 327;
Best Local Similarity 33.9%; Pred. No. 0.047;
Matches 20; Conservative 16; Mismatches 22; Indels 1; Gaps 1;

Qy 6 PRKRAKTEKEORRIERILNRRAAHQSKREKKRLHLOYLERKCSLLENLLNSVLEKL 64
Db 258 PALPTQPAEAPRKREVLNMRRAECRRKKYKVCLENRAVAVLEN-QNKTLEL 315

RESULT 7

US-09-252-658-6

Sequence 6, Application US/09252658
Patent No. 6251667

GENERAL INFORMATION:
APPLICANT: HABENER, JOEL F.
APPLICANT: HOFFLER, JAMES P.
TITLE OF INVENTION: A CAMP-RESPONSIVE TRANSCRIPTIONAL
TITLE OF INVENTION: ENHANCER BINDING PROTEIN
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
STREET: 1100 NEW YORK AVE., NW, STE. 600
CITY: WASHINGTON
STATE: DC
COUNTRY: USA
ZIP: 20005

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/252,658
FILING DATE: HEREWITH
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/684,965
FILING DATE: 22-MAY-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US89/05234
FILING DATE: 20-NOV-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/272,980
FILING DATE: 18-NOV-1988
ATTORNEY/AGENT INFORMATION:
NAME: Cimbala, Michele A
REGISTRATION NUMBER: 33,851
REFERENCE/DOCKET NUMBER: 0609.1990002
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202)371-2600
TELEFAX: (202)371-2540

INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 327 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein

```
/
/ FILING DATE: 31-JUL-1995
/ APPLICATION NUMBER: 60/018,496
/ FILING DATE: 29-MAY-1996
/ APPLICATION NUMBER: 08/690,011
/ FILING DATE: 31-JULY-1996
/ ATTORNEY/AGENT INFORMATION:
/   NAME: Serunian, Leslie A.
/   REGISTRATION NUMBER: 35,353
/   REFERENCE/DOCKET NUMBER: 2026-4199US2
/ TELECOMMUNICATION INFORMATION:
/   TELEPHONE: (212)758-4800
/   TELEFAX: (212)751-6849
/ INFORMATION FOR SEQ ID NO: 7:
/   SEQUENCE CHARACTERISTICS:
/     LENGTH: 68 amino acids
/     TYPE: amino acid
/     STRANDEDNESS: unknown
/     TOPOLOGY: linear
/     MOLECULE TYPE: peptide
/     SEQUENCE DESCRIPTION: SEQ ID NO: 7:
US-09-299-495F-7

Query Match          25.6%; Score 87.5; DB 2; Length 68;
Best Local Similarity 37.3%; Pred. No. 0.011;
Matches 19; Conservative 15; Mismatches 16; Indels 1; Gaps 1;

QY 14 EEKEQRRIERILNRRAAHQSKREKRLHLYLERKCSLLENLLNSVNLK 64
Db 7 EEARKREVRLMKNREAAECRRKKYVKCLENRVAVLEN-QNKTLIBEL 56

RESULT 10
US-08-690-011A-7
/ Sequence 7, Application US/08690011A
/ Patent No. 5942433
/ GENERAL INFORMATION:
/   APPLICANT: VINSON, Charles R.
/   TITLE OF INVENTION: EXTENSION OF A PROTEIN-PROTEIN
/   TITLE OF INVENTION: INTERACTION SURFACE TO INACTIVATE THE FUNCTION OF A
/   TITLE OF INVENTION: CELLULAR PROTEIN
/   NUMBER OF SEQUENCES: 60
/   CORRESPONDENCE ADDRESS:
/     ADDRESSEE: MORGAN & FINNEGAN, L.L.P.
/     STREET: 345 Park Avenue
/     CITY: New York
/     STATE: NY
/     COUNTRY: USA
/     ZIP: 10154-0053
/ COMPUTER READABLE FORM:
/   MEDIUM TYPE: Floppy disk
/   COMPUTER: IBM PC compatible
/   OPERATING SYSTEM: PC-DOS/MS-DOS
/   SOFTWARE: Patent In Release #1.0, Version #1.30B
/ CURRENT APPLICATION DATA:
/   APPLICATION NUMBER: US/08/690,011A
/   FILING DATE: 31-JUL-1996
/   PRIOR APPLICATION DATA:
/     APPLICATION NUMBER: 60/001,654
/     FILING DATE: 31-JUL-1995
/   PRIOR APPLICATION DATA:
/     APPLICATION NUMBER: 60/018,496
/     FILING DATE: 29-MAY-1996
/   ATTORNEY/AGENT INFORMATION:
/     NAME: Serunian, Leslie A.
/     REGISTRATION NUMBER: 35,353
/     REFERENCE/DOCKET NUMBER: 2026-4199US1
/   TELECOMMUNICATION INFORMATION:
/     TELEPHONE: (212)751-6849
/     TELEFAX: (212)751-6849
/ INFORMATION FOR SEQ ID NO: 7:
/   SEQUENCE CHARACTERISTICS:
/     LENGTH: 69 amino acids
```

```
/
/ TYPE: amino acid
/ STRANDEDNESS: unknown
/ TOPOLOGY: linear
/ MOLECULE TYPE: peptide
US-08-690-011A-7

Query Match          25.6%; Score 87.5; DB 1; Length 69;
Best Local Similarity 37.3%; Pred. No. 0.011;
Matches 19; Conservative 15; Mismatches 16; Indels 1; Gaps 1;

QY 14 EEKEQRRIERILNRRAAHQSKREKRLHLYLERKCSLLENLLNSVNLK 64
Db 7 EEARKREVRLMKNREAAECRRKKYVKCLENRVAVLEN-QNKTLIBEL 56

RESULT 11
US-09-299-495F-9
/ Sequence 9, Application US/09299495F
/ Patent No. 6361968
/ GENERAL INFORMATION:
/   APPLICANT: VINSON, Charles R.
/   TITLE OF INVENTION: EXTENSION OF A PROTEIN-PROTEIN
/   TITLE OF INVENTION: INTERACTION SURFACE TO INACTIVATE THE FUNCTION OF A
/   TITLE OF INVENTION: CELLULAR PROTEIN
/   NUMBER OF SEQUENCES: 64
/   CORRESPONDENCE ADDRESS:
/     ADDRESSEE: MORGAN & FINNEGAN, L.L.P.
/     STREET: 345 Park Avenue
/     CITY: New York
/     STATE: NY
/     COUNTRY: USA
/     ZIP: 10154-0053
/ COMPUTER READABLE FORM:
/   MEDIUM TYPE: Floppy disk
/   COMPUTER: IBM PC compatible
/   OPERATING SYSTEM: MS WORD 97
/   SOFTWARE: ASCII
/ CURRENT APPLICATION DATA:
/   APPLICATION NUMBER: US/09/299,495F
/   FILING DATE: 26-Apr-1999
/   PRIOR APPLICATION DATA:
/     APPLICATION NUMBER: 60/001,654
/     FILING DATE: 31-JUL-1995
/     APPLICATION NUMBER: 60/018,496
/     FILING DATE: 29-MAY-1996
/     APPLICATION NUMBER: 08/690,011
/     FILING DATE: 31-JULY-1996
/   ATTORNEY/AGENT INFORMATION:
/     NAME: Serunian, Leslie A.
/     REGISTRATION NUMBER: 35,353
/     REFERENCE/DOCKET NUMBER: 2026-4199US2
/   TELECOMMUNICATION INFORMATION:
/     TELEPHONE: (212)758-4800
/     TELEFAX: (212)751-6849
/ INFORMATION FOR SEQ ID NO: 9:
/   SEQUENCE CHARACTERISTICS:
/     LENGTH: 92 amino acids
/     TYPE: amino acid
/     STRANDEDNESS: unknown
/     TOPOLOGY: linear
/     MOLECULE TYPE: peptide
/     SEQUENCE DESCRIPTION: SEQ ID NO: 9:
US-09-299-495F-9

Query Match          25.6%; Score 87.5; DB 2; Length 92;
Best Local Similarity 37.3%; Pred. No. 0.015;
Matches 19; Conservative 15; Mismatches 16; Indels 1; Gaps 1;

QY 14 EEKEQRRIERILNRRAAHQSKREKRLHLYLERKCSLLENLLNSVNLK 64
Db 31 EEARKREVRLMKNREAAECRRKKYVKCLENRVAVLEN-QNKTLIBEL 80
```

```
RESULT 12
US-08-690-011A-9
; Sequence 9, Application US/08690011A
; Patent No. 5942433
; GENERAL INFORMATION:
; APPLICANT: VINSON, Charles R.
; APPLICANT: KRYLOV, Dmitry
; TITLE OF INVENTION: EXTENSION OF A PROTEIN-PROTEIN
; TITLE OF INVENTION: INTERACTION SURFACE TO INACTIVATE THE FUNCTION OF A
; TITLE OF INVENTION: CELLULAR PROTEIN
; NUMBER OF SEQUENCES: 60
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORGAN & FINNEGAN, L.L.P.
; STREET: 345 Park Avenue
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10154-0053
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30B
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/690,011A
; FILING DATE: 31-JUL-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/001,654
; FILING DATE: 31-JUL-1995
; APPLICATION DATA:
; APPLICATION NUMBER: 60/018,496
; FILING DATE: 29-MAY-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Serunian, Leslie A.
; REGISTRATION NUMBER: 35,353
; REFERENCE/DOCKET NUMBER: 2026-4199US1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212)758-4800
; TELEFAX: (212)751-6849
; INFORMATION FOR SEQ ID NO: 9:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 93 amino acids
; TYPE: amino acid
; STRANDEDNESS: unknown
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-690-011A-9

Query Match 25.6%; Score 87.5; DB 1; Length 93;
Best Local Similarity 37.3%; Pred. No. 0.015;
Matches 19; Conservative 15; Mismatches 16; Indels 1; Gaps 1;

QY 14 EEKEORRIERILNRRAAHQSRKKRLHLYLERKCSLLENLLNSVNLKL 64
Db 31 EEARKREVLNKNREARECRKKYVKLCLENRVAVLEN-QNKTLLIEL 80

RESULT 13
US-09-949-016-7832
; Sequence 7832, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11281
; LENGTH: 501
; TYPE: PRT
; ORGANISM: Human

Query Match 25.6%; Score 87.5; DB 1; Length 93;
Best Local Similarity 37.3%; Pred. No. 0.015;
Matches 19; Conservative 15; Mismatches 16; Indels 1; Gaps 1;

QY 14 EEKEORRIERILNRRAAHQSRKKRLHLYLERKCSLLENLLNSVNLKL 64
Db 31 EEARKREVLNKNREARECRKKYVKLCLENRVAVLEN-QNKTLLIEL 80

RESULT 14
US-09-640-211A-1046
; Sequence 1046, Application US/09640211A
; Patent No. 6833446
; GENERAL INFORMATION:
; APPLICANT: Wood, Marion
; APPLICANT: Shenk, Michael A.
; APPLICANT: McGrath, Annette
; APPLICANT: Glenn, Matthew
; TITLE OF INVENTION: Compositions and Methods for the
; TITLE OF INVENTION: Modification of Gene Transcription
; FILE REFERENCE: 11000.1021C1U
; CURRENT APPLICATION NUMBER: US/09/640,211A
; CURRENT FILING DATE: 2000-08-16
; NUMBER OF SEQ ID NOS: 2368
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1046
; LENGTH: 102
; TYPE: PRT
; ORGANISM: Pinus radiata
US-09-640-211A-1046

Query Match 25.1%; Score 86; DB 2; Length 102;
Best Local Similarity 40.8%; Pred. No. 0.025;
Matches 20; Conservative 9; Mismatches 20; Indels 0; Gaps 0;

QY 6 PKRAKTKKEEQRRIERILNRRAAHQSRKKRLHLYLERKCSLLEN 54
Db 39 PKRPRSDLSAEKKEARAHNRITAAQNSRDKRKQQTSLQORVIDLEN 87

RESULT 15
US-09-949-016-11281
; Sequence 11281, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11281
; LENGTH: 501
; TYPE: PRT
; ORGANISM: Human
```

```
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7832
; LENGTH: 315
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-7832

Query Match 25.3%; Score 86.5; DB 2; Length 315;
Best Local Similarity 37.3%; Pred. No. 0.076;
Matches 19; Conservative 15; Mismatches 16; Indels 1; Gaps 1;

QY 14 EEKEORRIERILNRRAAHQSRKKRLHLYLERKCSLLENLLNSVNLKL 64
Db 254 EEATRKREVLNKNREARECRKKYVKLCLENRVAVLEN-QNKTLLIEL 303

RESULT 14
US-09-640-211A-1046
; Sequence 1046, Application US/09640211A
; Patent No. 6833446
; GENERAL INFORMATION:
; APPLICANT: Wood, Marion
; APPLICANT: Shenk, Michael A.
; APPLICANT: McGrath, Annette
; APPLICANT: Glenn, Matthew
; TITLE OF INVENTION: Compositions and Methods for the
; TITLE OF INVENTION: Modification of Gene Transcription
; FILE REFERENCE: 11000.1021C1U
; CURRENT APPLICATION NUMBER: US/09/640,211A
; CURRENT FILING DATE: 2000-08-16
; NUMBER OF SEQ ID NOS: 2368
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1046
; LENGTH: 102
; TYPE: PRT
; ORGANISM: Pinus radiata
US-09-640-211A-1046

Query Match 25.1%; Score 86; DB 2; Length 102;
Best Local Similarity 40.8%; Pred. No. 0.025;
Matches 20; Conservative 9; Mismatches 20; Indels 0; Gaps 0;

QY 6 PKRAKTKKEEQRRIERILNRRAAHQSRKKRLHLYLERKCSLLEN 54
Db 39 PKRPRSDLSAEKKEARAHNRITAAQNSRDKRKQQTSLQORVIDLEN 87

RESULT 15
US-09-949-016-11281
; Sequence 11281, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 11281
; LENGTH: 501
; TYPE: PRT
; ORGANISM: Human
```

US-09-949-016-11281

Query Match 25.0%, Score 85.5; DB 2; Length 501;
Best Local Similarity 35.5%, Pred. No. 0.16;
Matches 22; Conservative 16; Mismatches 23; Indels 1; Gaps 1;

Qy 1 KSTLPPRRKAKTKEEKEQRIERILNRRAAHQSRKRLHLOYLERKCSLLENLLNSVN 60
Db 333 QSTSGRRFRANEDPDEKRR-KFLERNFAAASRCQKQKRWVQSLKKAEDLSSINGQLQ 391

Qy 61 LE 62

Db 392 SE 393

Search completed: November 23, 2005, 03:27:34
Job time : 7.65899 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 23, 2005, 03:05:10 ; Search time 29.2212 Seconds
(without alignments)
1641.819 Million cell updates/sec

Title: US-10-663-450-60

Perfect score: 342

Sequence: 1 KSTLPKRAKTKSEKQRR.....CSLLENLLNSVNLKLAHDE 68

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Uniprot_05.80.*

1: uniprot_sprot.*

2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	342	100.0	230	1	HAC1 YEAST	P41546 saccharomyc
2	260	76.0	228	2	Q75BQ5 ASHGO	Q75BQ5 ashbya gos
3	257	75.1	273	2	Q6CKQ1 KLJULA	Q6ckq1 kluyveromyc
4	214	62.6	357	2	Q5AA52 CANAL	Q5aa52 candida alb
5	211.5	61.8	260	2	Q6BQC2 DEBHA	Q6bqc2 debaryomyce
6	201	58.8	299	2	Q6CEV1 YARLI	Q6cev1 yarrowia li
7	196	57.3	451	2	Q8VFF3 TRIRE	Q8tf3 trichoderma
8	193	56.4	430	2	Q7SHF0 NEOCR	Q7shf0 neurospora
9	182	53.2	556	2	Q51KW8 MAGGR	Q51kw8 magnaporthe
10	176.5	51.6	342	2	Q6W8X2 ASPNG	Q6w8x2 aspergillus
11	176.5	51.6	347	2	Q5AQN3 EMENI	Q5agn3 aspergillus
12	176.5	51.6	350	2	Q8TFU8 EMENI	Q8tfu8 emericalia
13	176.5	51.6	433	2	Q4WEY8 ASPFU	Q4wey8 aspergillus
14	173	50.6	329	2	Q6FLY3 CANGA	Q6fly3 candida gla
15	157	45.9	429	2	Q4HTT5 GIBZE	Q4ht5 gibberella
16	124.5	36.4	168	1	H5Y ARATH	Q24646 arabidopsis
17	122	35.7	336	2	Q5VR11 ORYSA	Q5vr11 oryza sativ
18	121.5	35.5	69	2	Q5BL60 BRACM	Q5bl60 brassica ca
19	121.5	35.5	109	2	Q84XX8 BRARP	Q84xx8 brassica ra
20	110.5	32.3	322	2	Q04233 VICFA	Q04233 vicia faba
21	109.5	32.0	321	2	Q8GRY7 LOTJA	Q8gry7 lotus japon
22	108.5	31.7	176	2	Q69XK6 ORYSA	Q69xk6 oryza sativ
23	107	31.3	120	2	Q682B6 ARATH	Q682b6 arabidopsis
24	107	31.3	149	1	HYH ARATH	Q8w191 arabidopsis
25	106.5	31.1	158	1	HYE LYCES	Q9em50 lycopersico
26	106	31.0	188	2	Q6ZHT8 ORYSA	Q6zht8 oryza sativ
27	104.5	30.6	322	2	Q39896 SOYBN	Q39896 glycine max
28	104.5	30.6	326	2	Q39895 SOYBN	Q39895 glycine max
29	103	30.1	646	2	Q6AU90 ORYSA	Q6au90 oryza sativ
30	102.5	30.0	141	2	Q677A7 ORYSA	Q677a7 oryza sativ
31	96	28.1	267	2	Q7FTU1 ANOGA	Q7ftj1 anopheles 9

RESULT 1

HAC1_YEAST	ID	HAC1_YEAST	STANDARD;	PRT;	230 AA.
AC	P41546;	P87040;			
DT	01-NOV-1995	(Rel. 32, Created)			
DT	01-NOV-1995	(Rel. 32, Last sequence update)			
DT	13-SEP-2005	(Rel. 48, Last annotation update)			
DE	HAC1	protein.			
GN	Name=HAC1; Synonyms=ERN4, IRE2; OrderedLocusNames=YPL031W;				
OS	Saccharomyces cerevisiae (Baker's Yeast).				
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;				
OC	Saccharomycetales; Saccharomycetaceae; Saccharomycetes.				
OX	NCBI_TaxID=4932;				
RN	[1]				
RP	NUCLEOTIDE SEQUENCE.				
RC	STRAIN=S288c;				
RX	MEDLINE=95116316; PubMed=7816617;				
RA	Nojima H., Leem S.-H., Araki H., Sakai A., Nakashima N., Kanaoka Y.,				
RA	Ono Y.;				
RT	"Hac1: a novel yeast bZIP protein binding to the CRE motif is a				
RT	multicopy suppressor for cdc10 mutant of Schizosaccharomyces pombe."				
RL	Nucleic Acids Res. 22:5279-5288(1994).				
RN	[2]				
RP	NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].				
RC	STRAIN=S288c / AB972;				
RX	MEDLINE=95400292; PubMed=7670463;				
RA	Murakami Y., Naitou M., Hagiwara H., Shibata T., Ozawa M.,				
RA	Sasanuma Y.-I., Sasanuma M., Teuchiya Y., Soeda E., Yokoyama K.,				
RA	Yamazaki M.-A., Tashiro H., Eki T.;				
RT	"Analysis of the nucleotide sequence of chromosome VI from				
RT	Saccharomyces cerevisiae."				
RL	Nat. Genet. 10:261-268(1995).				
RN	[3]				
RP	SEQUENCE REVISION TO 183-230.				
RA	Murakami Y.;				
RL	Submitted (JAN-1998) to the EMBL/GenBank/DBJ databases.				
RN	[4]				
RP	NUCLEOTIDE SEQUENCE.				
RC	STRAIN=AB320;				
RX	MEDLINE=97222447; PubMed=9077435;				
RA	Mori K., Kawahara T., Yoshida H., Yanagi H., Yura T.;				
RT	"Signalling from endoplasmic reticulum to nucleus: transcription				
RT	factor with a basic-leucine zipper motif is required for the unfolded				
RT	protein-response pathway."				
RL	Genes Cells 1:803-817(1996).				
RN	[5]				
RP	CHARACTERIZATION.				
RX	MEDLINE=97086687; PubMed=8932376; DOI=10.1093/nar/24.21.4222;				
RA	Nikawa J.-I., Akiyoshi M., Hirata S., Fukuda T.;				
RT	"Saccharomyces cerevisiae IRE2/HAC1 is involved in IRE1-mediated KAR2				
RT	expression."				
RL	Nucleic Acids Res. 24:4222-4226(1996).				
RN	[6]				
RP	CHARACTERIZATION, AND ALTERNATIVE SPLICING.				

32	95.5	27.9	786	2	Q86AS2_DICDI	Q86as2 dictyosteli
33	95.5	27.9	787	2	Q554P0_DICDI	Q554p0 dictyosteli
34	95	27.8	372	2	Q54R29_DICDI	Q54r29 dictyosteli
35	92.5	27.0	327	2	Q6NVP9_XENTR	Q6nvp9 xenopus tro
36	92.5	27.0	703	2	Q4WC74_ASPFU	Q4wc74 aspergillus
37	91.5	26.8	306	2	Q81ON6_CABEL	Q81on6 caenorhabdi
38	91.5	26.8	314	2	Q814D8_CABEL	Q814d8 caenorhabdi
39	91.5	26.8	324	2	Q8AVY5_XENLA	Q8avy5 xenopus lae
40	91.5	26.8	325	2	Q9U211_CABEL	Q9u211 caenorhabdi
41	91.5	26.8	327	2	Q9U210_CABEL	Q9u210 caenorhabdi
42	91.5	26.8	331	2	Q814D7_CABEL	Q814d7 caenorhabdi
43	91.5	26.8	333	2	Q61BR7_CABBR	Q61br7 caenorhabdi
44	90.5	26.5	568	2	Q8LIB3_ORYSA	Q8lib3 oryza sativ
45	90	26.3	627	2	Q5BD44_EMENI	Q5bd44 aspergillus

ALIGNMENTS

```

RX MEDLINE=97053779; PubMed=8898193; DOI=10.1016/S0092-8674(00)81360-4;
RA Cox J.S., Walter P.;
RT "A novel mechanism for regulating activity of a transcription factor
RT that controls the unfolded protein response.";
RL Cell 87:1391-404(1996).
CC -!- FUNCTION: Seems to be involved in the unfolded protein response
CC (UPR) pathway. Binds to the UPR element (UPRE) in the promoter of
CC UPR-regulated genes such as KAR2, PD11, EUG1 and PKB2; activates
CC the transcription of these genes.
CC -!- SUBUNIT: Homodimer.
CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).
CC -!- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=2;
CC Comment=The level of each isoform is regulated by a splicing
CC event that occurs when the UPR is induced by IRE1;
CC Name=U;
CC isoId=P41546-1; Sequences=Displayed;
CC Name=I;
CC isoId=P41546-2; Sequences=VSP_000586;
CC Notes=Active and stable isoform which induces UPR;
CC -!- SIMILARITY: Belongs to the bZIP family.
CC -!- SIMILARITY: Contains 1 bZIP domain.
CC -----
CC This Swiss-Prot entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use as long as its content is in no way modified and this statement is not
CC removed.
CC -----
DR EMBL; D26506; BAA05513.1; -; Genomic DNA.
DR EMBL; D50617; BAA24425.1; -; Genomic DNA.
DR EMBL; D86413; BAA19565.1; -; Genomic DNA.
DR PIR; S78571; S78571.
DR GeneOnline; 140124; -.
DR TRANSFAC; T02039; -.
DR Ensemble; YFL031W; Saccharomyces cerevisiae.
DR SGD; S00001863; HAC1.
DR GO; GO:0005634; C:nucleus; TAS.
DR GO; GO:0003704; F:specific RNA polymerase II transcription fa. .; TAS.
DR GO; GO:0003700; F:transcription factor activity; TAS.
DR GO; GO:0005644; P:phospholipid metabolism; TAS.
DR GO; GO:0045893; P:positive regulation of transcription, DNA-d. .; IDA.
DR GO; GO:0006357; P:regulation of transcription from RNA polyme. .; TAS.
DR GO; GO:0006990; P:unfolded protein response, positive regulat. .; TAS.
DR InterPro; IPR011700; bZIP_2
DR Pfam; PF07716; bZIP_2; 1.
DR SMART; SM00338; BRLZ; 1.
DR PROSITE; PS00217; bZIP; FALSE NEG.
DR PROSITE; PS00036; bZIP_BASIC; 1.
KW Activator; Alternative_splicing; Complete proteome; DNA-binding;
KW Nuclear protein; Transcription; Transcription regulation;
KW Unfolded protein response.
FT DOMAIN 67 81
FT DNA_BIND 29 61
FT VARSPLIC 221 230
FT FTID=VSP_000586.
FT RLCRPRVCFRVGRDFMGAAECLRRKMYQSRRLPYVTIN
FT NLFDAVASPLADPLCDDIAGNSLPFDNSIDLNRWRNPVIT
FT MKTRKQ -> ATLSPKSMRDSASDQETSWELOMFKTENPVE
FT STTLPADVNNLFDVASRWQTHSATI (in Ref. 2).
FT RLCRPRVCFRVGRDFMGAAECLRRKMYQSRRLPYVTI
FT -> ATLSPKSMRDSASDQETSWELOMFKTENPVESTLTPAV
FT DN (in Ref. 4).
SQ SEQUENCE 230 AA; 26582 MW; 43073BCCCC4709B CRC64;
Query Match 100.0%; Score 342; DB 1; Length 230;
Best Local Similarity 100.0%; Pred. No. 2,3e-23;
Matches 68; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 KSTLPPRKAKTKKEQRIIRLNRAAHQSRKRLHLQYLERKCSLLENLNSVN 60
|||||

```

```

Db 23 KSTLPPRKAKTKKEQRIIRLNRAAHQSRKRLHLQYLERKCSLLENLNSVN 82
QY 61 LEKLADHE 68
Db 83 LEKLADHE 90
RESULT 2
Q75B05 ASHGO
ID Q75B05_ASHGO PRELIMINARY; PRT; 228 AA.
AC Q75B05;
DT 05-JUL-2004 (TReMBLrel. 27, Created)
DT 05-JUL-2004 (TReMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TReMBLrel. 27, Last annotation update)
DE ACR216Cp.
GN Name=ACR216C;
OS Ashbya gossypii (Yeast) (Eremothecium gossypii).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Eremothecium.
OX NCBI_TaxID=33169;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=ATCC 10895;
RX PubMed=15001715; DOI=10.1126/science.1095781;
RA Dietrich F.S., Voegelé S., Brachat S., Lerch A., Gates K., Steiner S.,
RA Mohr C., Poehmann R., Luedi P., Choi S., Wing R.A., Flavien A.,
RA Gaffney T.D., Philippsen P.;
RT "The Ashbya gossypii genome as a tool for mapping the ancient
RT Saccharomyces cerevisiae genome.";
RL Science 304:304-307(2004).
DR EMBL; AE016816; AAS51442.1; -; Genomic_DNA.
DR AGD; ACR216C; -.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR011700; bZIP_2.
DR Pfam; PF07716; bZIP_2; 1.
DR PROSITE; PS00036; bZIP_BASIC; 1.
KW Complete proteome.
SQ SEQUENCE 228 AA; 24914 MW; 605A60D78633F038 CRC64;
Query Match 76.0%; Score 260; DB 2; Length 228;
Best Local Similarity 81.2%; Pred. No. 6.7e-16;
Matches 52; Conservative 6; Mismatches 6; Indels 0; Gaps 0;
QY 1 KSTLPPRKAKTKKEQRIIRLNRAAHQSRKRLHLQYLERKCSLLENLNSVN 60
Db 14 KSTLPPRKAKTKKEQRIIRLNRAAHQSRKRLHLQYLERKCSLLENLNSVN 73
QY 61 LEKL 64
Db 74 LGAL 77
RESULT 3
Q6CKQ1 KLULA
ID Q6CKQ1_KLULA PRELIMINARY; PRT; 273 AA.
AC Q6CKQ1;
DT 25-OCT-2004 (TReMBLrel. 28, Created)
DT 25-OCT-2004 (TReMBLrel. 28, Last sequence update)
DT 25-OCT-2004 (TReMBLrel. 28, Last annotation update)
DE Kluyveromyces lactis strain NRRL Y-1140 chromosome F of strain NRRL Y-
DE 1140 of Kluyveromyces lactis.
GN OrderedLocustNames=KLUA0F08976g;
OS Kluyveromyces lactis (Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Kluyveromyces.
OX NCBI_TaxID=28985;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=CBS 2359 / IFO 1267 / NRRL Y-1140 / WM37;
RX PubMed=15229592; DOI=10.1038/nature02579;

```



```

RA Dujon B., Sherman D., Fischer G., Durrens P., Casaregola S.,
RA Lafontaine I., de Montigny J., Marek C., Neuveglise C., Talla E.,
RA Goffard N., Frangeul L., Aigle M., Anthouard V., Babour A., Barbe V.,
RA Barnay S., Blanchin S., Beckerich J.-M., Beyne E., Bleykasten C.,
RA Boisrame A., Boyer J., Cattolico L., Confanioli F., de Daruvar A.,
RA Despons L., Fabre E., Fairhead C., Ferry-Dumazet H., Groppi A.,
RA Hantraye F., Hennequin C., Jauniaux N., Joyet P., Kachouri R.,
RA Kerrest A., Koszul R., Lemaire M., Lesur I., Ma L., Muller H.,
RA Nicaud J.-M., Nikolski M., Oztas S., Ozier-Kalogeropoulos O.,
RA Pellenz S., Potier S., Richard G.-P., Straub M.-L., Suleau A.,
RA Swennen D., Tekala F., Wesolowski-Louvel M., Westhof E., Wirth B.,
RA Zeniou-Meyer M., Zivanovic Y., Bolotin-Fukuhara M., Thierry A.,
RA Bouchier C., Caudron B., Scarpelli C., Gaillardin C., Weissenbach J.,
RA Wincker P., Souciet J.-L.;
RT "Genome evolution in Yeasts.";
RL Nature 430:35-44 (2004).
DR EMBL; CR382126; CAG98196.1; -; Genomic_DNA.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR011616; bZIP_1.
DR Pfam; PF001170; bZIP_1; 1.
DR SMART; SM00338; BRLZ; 1.
DR PROSITE; PS00036; bZIP BASIC; 1.
KW Complete proteome; DNA-binding; Nuclear protein.
SQ SEQUENCE 273 AA; 30429 MW; 97DFDF7B1F8ACF5 CRC64;

Query Match 75.1%; Score 257; DB 2; Length 273;
Best Local Similarity 76.1%; Pred. No. 1.5e-15;
Matches 51; Conservative 8; Mismatches 8; Indels 0; Gaps 0;

QY 1 KSTLPPRKRAKTKKEEQRIERILNRRRAAHSREKRLHLQYLERKCSLLENLNSVN 60
DB 15 KPTLPPRKRAKTKKEEQRIERILNRRRAAHSREKRLHLQYLERKCSLLENLNSVN 74
QY 61 LEKLADH 67
DB 75 LDILSEN 81

RESULT 4
Q5AA52 CANAL PRELIMINARY; PRT; 357 AA.
AC Q5AA52;
DT 10-MAY-2005 (TrEMBLrel. 30, Created)
DT 10-MAY-2005 (TrEMBLrel. 30, Last sequence update)
DE Hypothetical protein HAC1.
GN Name=HAC1; ORFNames=CaO19.2432, CaO19.9968;
OS Candida albicans SC5314.
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Mitosporic Saccharomycetales; Candida.
OX NCBI_TaxID=237561;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC PubMed=15123810; DOI=10.1073/pnas.0401648101;
RX Jones T., Federapfel N.A., Chibana H., Dungan J., Kalman S.,
RA Magee B.B., Newport G., Thorstenson Y.R., Agabian N., Magee P.T.,
RA Davis R.W., Scherer S.;
RT "The diploid genome sequence of Candida albicans.";
RL Proc. Natl. Acad. Sci. U.S.A. 101:7329-7334 (2004).
RN [2]
RP NUCLEOTIDE SEQUENCE.
RC Dungan J., Kuo A., Newport G., Ian C.-Y., Iijima C., Adegbola O.,
RA Roberts J., Persson K., Donnelly S., Favoreto S., Tzung K.-W.,
RA Jones T., Scherer S., Agabian N.;
RT "Annotation of the Genome of Candida albicans.";
RL Submitted (APR-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; AACQ0100041; EAK99529.1; -; Genomic DNA.
DR EMBL; AACQ0100040; EAK99617.1; -; Genomic DNA.

```

```

DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
KW DNA-binding; Hypothetical protein; Nuclear protein.
SQ SEQUENCE 357 AA; 40127 MW; 8E1C633A2B4ECDAE CRC64;

Query Match 62.6%; Score 214; DB 2; Length 357;
Best Local Similarity 73.3%; Pred. No. 1.6e-11;
Matches 44; Conservative 5; Mismatches 11; Indels 0; Gaps 0;

QY 1 KSTLPPRKRAKTKKEEQRIERILNRRRAAHSREKRLHLQYLERKCSLLENLNSVN 60
DB 54 KSTLPPRKRAKTKKEEQRIERILNRRRAAHSREKRLHLQYLERKCSLLENLNSVN 113

RESULT 5
Q6BQC2 DEBHA PRELIMINARY; PRT; 260 AA.
AC Q6BQC2;
DT 25-OCT-2004 (TrEMBLrel. 28, Created)
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE Similarities with sp|P41546 Saccharomyces cerevisiae HAC1 protein.
GN OrderedLocusNames=DEHA0E07139g;
OS Debaryomyces hansenii (Yeast) (Torulaspora hansenii).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Debaryomycetes.
OX NCBI_TaxID=4959;
RN [1]
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=ATCC 36239 / CBS 767;
RX PubMed=15295992; DOI=10.1038/nature02579;
RA Dujon B., Sherman D., Fischer G., Durrens P., Casaregola S.,
RA Lafontaine I., de Montigny J., Marek C., Neuveglise C., Talla E.,
RA Goffard N., Frangeul L., Aigle M., Anthouard V., Babour A., Barbe V.,
RA Barnay S., Blanchin S., Beckerich J.-M., Beyne E., Bleykasten C.,
RA Boisrame A., Boyer J., Cattolico L., Confanioli F., de Daruvar A.,
RA Despons L., Fabre E., Fairhead C., Ferry-Dumazet H., Groppi A.,
RA Hantraye F., Hennequin C., Jauniaux N., Joyet P., Kachouri R.,
RA Kerrest A., Koszul R., Lemaire M., Lesur I., Ma L., Muller H.,
RA Nicaud J.-M., Nikolski M., Oztas S., Ozier-Kalogeropoulos O.,
RA Pellenz S., Potier S., Richard G.-P., Straub M.-L., Suleau A.,
RA Swennen D., Tekala F., Wesolowski-Louvel M., Westhof E., Wirth B.,
RA Zeniou-Meyer M., Zivanovic Y., Bolotin-Fukuhara M., Thierry A.,
RA Bouchier C., Caudron B., Scarpelli C., Gaillardin C., Weissenbach J.,
RA Wincker P., Souciet J.-L.;
RT "Genome evolution in Yeasts.";
RL Nature 430:35-44 (2004).
DR EMBL; CR382137; CAG87828.1; -; Genomic_DNA.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR011700; bZIP_2.
DR Pfam; PF07716; bZIP_2; 1.
DR SMART; SM00338; BRLZ; 1.
DR PROSITE; PS00217; bZIP; 1.
DR PROSITE; PS00036; bZIP BASIC; UNKNOWN 1.
KW Complete proteome; DNA-binding; Nuclear protein.
SQ SEQUENCE 260 AA; 29442 MW; 49B1819E393BFF6 CRC64;

Query Match 61.8%; Score 211.5; DB 2; Length 260;
Best Local Similarity 68.8%; Pred. No. 2e-11;
Matches 44; Conservative 7; Mismatches 8; Indels 5; Gaps 1;

QY 1 KSTLPPRKRAKTKKEEQRIERILNRRRAAHSREKRLHLQYLERKCSLLENLNSVN 60
DB 14 KSTLPPRKRAKTKKEEQRIERILNRRRAAHSREKRLHLQYLERKCSLLENLNSVN 68
QY 61 LEKL 64
DB 69 MSRL 72

```

```

RA Salotheimo M.L.A., Valkonen M., Penttilae M.E.;
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AJ413272; CAC98374.1; -: Genomic_DNA.
DR GO: GO:0005634; C:nucleus; IEA.
DR GO: GO:0003677; F:DNA binding; IEA.
DR GO: GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro: IPR011700; bZIP 2.
DR InterPro: IPR004827; TF bZIP.
DR Pfam: PF07716; bZIP 2; 1.
DR SMART: SM00338; BRLZ; 1.
DR PROSITE: PS50217; BZIP; 1.
DR KQ DNA-binding; Nuclear protein.
SQ SEQUENCE 451 AA; 49277 MW; 600F10E471EA3AD3 CRC64;

Query Match 57.3%; Score 196; DB 2; Length 451;
Best Local Similarity 61.9%; Pred. No. 8.6e-10;
Matches 39; Conservative 13; Mismatches 9; Indels 2; Gaps 1;

QY 1 KSTLPPRKRAKTKERKEQRRIILNRRAAHQSRKRLHLQYLERKCSLENLINSVN 60
Db : |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 102 KTNLPPRKRAKTKERKEQRRIILNRRAAHQSRKRLHLQYLERKCSLENLINSVN 159
Db : |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 61 LEK 63
Db 160 VQK 162

RESULT 8
QYSHFO_NEUCR PRELIMINARY; PRT; 430 AA.
ID Q7SHFO;
AC Q7SHFO;
DT 01-MAR-2004 (TrEMBLrel. 26, Created)
DT 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Predicted protein.
GN Name=NCU01856.1;
OS Neurospora crassa.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.
OX NCBI_TaxID=5141;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=OR74A;
RA Galagan J.E., Calvo S.E., Borkovich K.A., Selker E.U., Read N.D.,
RA Jaffe D., FitzHugh W., Ma L.-J., Smirnov S., Purcell S., Rehman B.,
RA Elkins T., Engels R., Wang S., Nielsen C.B., Butler J., Endrizzi M.,
RA Qui D., Ianakiev P., Pedersen D., Nelson M., Washburne M.,
RA Selitrennikoff C.P., Kinsey J.A., Braun E.L., Zelter A., Schulte U.,
RA Kothe G.O., Jedd G., Mewes W., Staben C., Marcotte E., Greenberg D.,
RA Roy A., Foley K., Naylor J., Thomann N., Barrett R., Gnerre S.,
RA Kamal M., Kamvysselis M., Mauceli E., Bielke C., Rudd S., Frishman D.,
RA Krystofova S., Rasmussen C., Metzberg R.L., Perkins D.D., Kroken S.,
RA Cogoni C., Macino G., Catchside D., Li W., Pratt R.J., Gemani S.A.,
RA DeSouza C.C., Glass L., Orbach M.J., Berglund J., Voelker R.,
RA Varian O., Plamann M., Seiler S., Dunlap J., Radford A., Aramayo R.,
RA Natvig D.O., Alex L.A., Mannhaupt G., Ebbold D.J., Freitag M.,
RA Paulsen I., Sachs M.S., Lander E.S., Nuebaum C., Birren B.;
RT "The Genome Sequence of the Filamentous Fungus Neurospora crassa.";
RL Nature 0:0-0(2003).
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL: AABX01000004; EAA36251.1; -: Genomic_DNA.
DR GO: GO:0005634; C:nucleus; IEA.
DR GO: GO:0003677; F:DNA binding; IEA.
DR GO: GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro: IPR011700; bZIP 2.
DR InterPro: IPR004827; TF bZIP.
DR Pfam: PF07716; bZIP 2; 1.
DR PROSITE: PS50217; BZIP; 1.
DR KQ DNA-binding; Nuclear protein.
SQ SEQUENCE 430 AA; 46599 MW; 0EB0657CFA6160DA CRC64;

Query Match 56.4%; Score 193; DB 2; Length 430;

RA Salotheimo M.L.A., Valkonen M., Penttilae M.E.;
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AJ413272; CAC98374.1; -: Genomic_DNA.
DR GO: GO:0005634; C:nucleus; IEA.
DR GO: GO:0003677; F:DNA binding; IEA.
DR GO: GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro: IPR011700; bZIP 2.
DR InterPro: IPR004827; TF bZIP.
DR Pfam: PF07716; bZIP 2; 1.
DR PROSITE: PS50217; BZIP; 1.
DR KQ DNA-binding; Nuclear protein.
SQ SEQUENCE 451 AA; 49277 MW; 600F10E471EA3AD3 CRC64;

Query Match 57.3%; Score 196; DB 2; Length 451;
Best Local Similarity 61.9%; Pred. No. 8.6e-10;
Matches 39; Conservative 13; Mismatches 9; Indels 2; Gaps 1;

QY 1 KSTLPPRKRAKTKERKEQRRIILNRRAAHQSRKRLHLQYLERKCSLENLINSVN 60
Db : |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 102 KTNLPPRKRAKTKERKEQRRIILNRRAAHQSRKRLHLQYLERKCSLENLINSVN 159
Db : |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 61 LEK 63
Db 160 VQK 162

RESULT 8
QYSHFO_NEUCR PRELIMINARY; PRT; 430 AA.
ID Q7SHFO;
AC Q7SHFO;
DT 01-MAR-2004 (TrEMBLrel. 26, Created)
DT 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Predicted protein.
GN Name=NCU01856.1;
OS Neurospora crassa.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.
OX NCBI_TaxID=5141;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=OR74A;
RA Galagan J.E., Calvo S.E., Borkovich K.A., Selker E.U., Read N.D.,
RA Jaffe D., FitzHugh W., Ma L.-J., Smirnov S., Purcell S., Rehman B.,
RA Elkins T., Engels R., Wang S., Nielsen C.B., Butler J., Endrizzi M.,
RA Qui D., Ianakiev P., Pedersen D., Nelson M., Washburne M.,
RA Selitrennikoff C.P., Kinsey J.A., Braun E.L., Zelter A., Schulte U.,
RA Kothe G.O., Jedd G., Mewes W., Staben C., Marcotte E., Greenberg D.,
RA Roy A., Foley K., Naylor J., Thomann N., Barrett R., Gnerre S.,
RA Kamal M., Kamvysselis M., Mauceli E., Bielke C., Rudd S., Frishman D.,
RA Krystofova S., Rasmussen C., Metzberg R.L., Perkins D.D., Kroken S.,
RA Cogoni C., Macino G., Catchside D., Li W., Pratt R.J., Gemani S.A.,
RA DeSouza C.C., Glass L., Orbach M.J., Berglund J., Voelker R.,
RA Varian O., Plamann M., Seiler S., Dunlap J., Radford A., Aramayo R.,
RA Natvig D.O., Alex L.A., Mannhaupt G., Ebbold D.J., Freitag M.,
RA Paulsen I., Sachs M.S., Lander E.S., Nuebaum C., Birren B.;
RT "The Genome Sequence of the Filamentous Fungus Neurospora crassa.";
RL Nature 0:0-0(2003).
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL: AABX01000004; EAA36251.1; -: Genomic_DNA.
DR GO: GO:0005634; C:nucleus; IEA.
DR GO: GO:0003677; F:DNA binding; IEA.
DR GO: GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro: IPR011700; bZIP 2.
DR InterPro: IPR004827; TF bZIP.
DR Pfam: PF07716; bZIP 2; 1.
DR PROSITE: PS50217; BZIP; 1.
DR SMART: SM00338; BRLZ; 1.
DR PROSITE: PS00036; BZIP_BASIC; UNKNOWN 1.
KW Complete proteome; DNA-binding; Nuclear protein.
SQ SEQUENCE 299 AA; 32817 MW; F526110CFB23ABB2 CRC64;

Query Match 58.8%; Score 201; DB 2; Length 299;
Best Local Similarity 70.2%; Pred. No. 2e-10;
Matches 40; Conservative 8; Mismatches 9; Indels 0; Gaps 0;

QY 4 LPPRKRAKTKERKEQRRIILNRRAAHQSRKRLHLQYLERKCSLENLINSVN 60
Db : |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 47 LPPRKRAKTKERKEQRRIILNRRAAHQSRKRLHLQYLERKCSLENLINSVN 103
Db : |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

RESULT 7
Q8TF33_TRIPE PRELIMINARY; PRT; 451 AA.
ID Q8TF33;
AC Q8TF33;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Transcription factor.
GN Name=hac1;
OS Trichoderma reesei (Hypocrea jecorina).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Hypocremycetidae; Hypocreales; Hypocreaceae; Hypocrea.
OX NCBI_TaxID=51453;
RN [1]
RP NUCLEOTIDE SEQUENCE.

```

Best Local Similarity 56.2%; Pred. No. 1.5e-09;
Matches 41; Conservative 12; Mismatches 14; Indels 6; Gaps 1;

Qy 1 KSTLPPRKAKTKEEQRRIERILNRRRAAQSREKKRLHLQYLRKCSLLENLNVN 56
105 KTNLPPRKAKTDEKEQRVERVLNRRRAAQSRRERKRLVEGLRRNKLETLMMQAQ 164
57 --NSVNLKLAHDH 67
165 QINQTLQALREN 177

Db

RESULT 9
Q51KW8 MAGGR
ID Q51KW8 MAGGR PRELIMINARY; PRT; 556 AA.
AC Q51KW8;
DT 13-SEP-2005 (T-EMBLrel. 31, Created)
DT 13-SEP-2005 (T-EMBLrel. 31, Last sequence update)
DT 13-SEP-2005 (T-EMBLrel. 31, Last annotation update)
DE Hypothetical protein.
GN ORFNames=MG09010.4;
OS Magnaporthe grisea 70-15.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariomycetes incertae sedis; Magnaportheaceae; Magnaporthe.
OX NCBI_TaxID=242507;
[1]
RN NUCLEOTIDE SEQUENCE.
RP STRAIN=70-15;
RA Birren B., Nuebaum C., Abebe A., Abouelleil A., Adekoya E.,
RA Alt-zahra M., Allen N., Allen T., An P., Anderson M., Anderson S.,
RA Arachchi H., Ambruster J., Bachantsang P., Baldwin J., Barry A.,
RA Bayul T., Blithsteyn B., Bloom T., Blye J., Boguialavsky L.,
RA Borowsky M., Boukgalter B., Brunsche A., Butler J., Calixte N.,
RA Calvo S., Camarata J., Campo K., Chang J., Cheshatsang Y., Citroen M.,
RA Collymore A., Considine T., Cook A., Cooke P., Corum B., Cuomo C.,
RA David R., Dawoe T., Degray S., Dodge S., Dooley K., Dorje P.,
RA Dorjee K., Dorris L., Duffey N., Dupes A., Elkins T., Engels R.,
RA Erickson J., Farina A., Faro S., Ferreira P., Fischer H.,
RA Fitzgerald M., Foley K., Gage D., Galagan J., Gearin G., Gnerre S.,
RA Gnikre A., Govette A., Graham J., Grandbois E., Gyaltsen K., Hafez N.,
RA Hagopian D., Hagos B., Hall J., Hatcher B., Heller A., Higgins H.,
RA Honan T., Horn A., Houde N., Hughes L., Hulme W., Husby E., Iliev I.,
RA Jaffe D., Jones C., Kamal M., Kamat A., Kamysseleis M., Karlsson E.,
RA Kells C., Kieu A., Klesner P., Kodira C., Kulbokas E., Labutti K.,
RA Lama D., Landers T., Leger J., Levine S., Lewis D., Lewis T.,
RA Lindblad-toh K., Liu X., Lokitsang T., Lokitsang Y., Lucien O.,
RA Lui A., Ma L.J., Mabbitt R., Macdonald J., Maclean C., Major J.,
RA Manning J., Marabelli R., Maru K., Matthews C., Maucelli E.,
RA McCarthy M., McDonough S., Mcghee T., Meldrum J., Meneus L.,
RA Mesirov J., Mihalev A., Mihova T., Mikkelson T., Menga V., Moru K.,
RA Mezes J., Multan L., Munson G., Naylor J., Neves C., Nguyen C.,
RA Nguyen N., Nguyen T., Nicol R., Nielsen C., Nizzari M., Norbu C.,
RA Norbu N., O'donnell P., Okaawa O., O'leary S., Omotosho B.,
RA O'Neill K., Oman S., Parker S., Perrin D., Phunkhang P., Piquani B.,
RA Purcell S., Rachupka T., Ramasamy U., Rameau R., Ray V., Raymond C.,
RA Retta R., Richardson S., Rise C., Rodriguez J., Rogers J., Rogov P.,
RA Rutman M., Schupbach R., Seaman C., Settipalli S., Sharpe T.,
RA Sheridan J., Sherpa N., Shi J., Smirnov S., Smith C., Sougnez C.,
RA Spencer B., Stalker J., Stange-thomson N., Stavropoulos S.,
RA Stetson K., Stone C., Stone S., Stubbs M., Talamas J., Tchuinga P.,
RA Tenzing P., Teefaye S., Theodore J., Thoulitsang Y., Topham K.,
RA Towey S., Tsamla T., Tsomo N., Vallee D., Vassiliev H.,
RA Venkataraman V., Vinson J., Vo A., Wade C., Wang S., Wangchuk T.,
RA Wangdi T., Whittaker C., Wilkinson J., Wu Y., Wyman D., Yadav S.,
RA Yang S., Yang X., Yeager S., Yee E., Young G., Zainoun J., Zembeck L.,
RA Zimmer A., Zody M., Lander E.;
RT "The genome sequence of Magnaporthe grisea."
RL Submitted (OCT-2003) to the EMBL/GenBank/DBJ databases.
[2]
RN NUCLEOTIDE SEQUENCE.
RP STRAIN=70-15;
RC Dean R., Mitchell T., Brown D., Pan H., Thon M.;
RA LBL Submitted (OCT-2003) to the EMBL/GenBank/DBJ databases.

Best Local Similarity 57.8%; Pred. No. 2e-08;
Matches 37; Conservative 14; Mismatches 11; Indels 2; Gaps 1;

Qy 1 KSTLPPRKAKTKEEQRRIERILNRRRAAQSREKKRLHLQYLRKCSLLENLNVN 60
105 KTNLPPRKAKTDEKEQRVERVLNRRRAAQSRRERKRLVEGLRRNKLETLMMQAQ 162
61 LEKL 64
163 FQKL 166

Db

RESULT 10
Q6W8X2 ASPNG
ID Q6W8X2 ASPNG PRELIMINARY; PRT; 342 AA.
AC Q6W8X2;
DT 05-JUL-2004 (T-EMBLrel. 27, Created)
DT 05-JUL-2004 (T-EMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (T-EMBLrel. 27, Last annotation update)
DE Transcription factor HACA.
GN Name=haca;
OS Aspergillus niger.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
OX NCBI_TaxID=5061;
[1]
RN NUCLEOTIDE SEQUENCE.
RP Mulder H.J.;
RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY303684; AAQ73495.1; -, Genomic_DNA.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003677; P:DNA binding; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR011700; BZIP_2.
DR InterPro; IPR004827; TF_BZIP.
DR Pfam; PF0716; BZIP_2; 1.
DR SMART; SM00338; BRLZ; 1.
DR PROSITE; PS0217; BZIP; 1.
DR PROSITE; PS00036; BZIP_BASIC; UNKNOWN_1.
KW DNA-binding; Nuclear protein.
SQ SEQUENCE 342 AA; 37148 MW; 623B3941A55C9C05 CRC64;

Query Match 51.6%; Score 176.5; DB 2; Length 342;
Best Local Similarity 55.9%; Pred. No. 3.9e-08;
Matches 38; Conservative 12; Mismatches 17; Indels 1; Gaps 1;

Qy 1 KSTLPPRKAKTKEEQRRIERILNRRRAAQSREKKRLHLQYLRKCSLLENLNVN 60
64 KTNLPPRKAKTDEKEQRRIERVLNRRRAAQSRRERKRLVEGLRRNKLETLMMQAQ 122
61 LEKLADHE 68
123 LQRLSQME 130

Db

Sat Nov 26 13:50:41 2005

```
RESULT 11
Q5AQN3_EMENI PRELIMINARY; PRT; 347 AA.
AC Q5AQN3;
DT 10-MAY-2005 (TrEMBLrel. 30, Created)
DT 10-MAY-2005 (TrEMBLrel. 30, Last sequence update)
DT 10-MAY-2005 (TrEMBLrel. 30, Last annotation update)
DE Hypothetical protein.
GN ORFNames=AN9397-2;
OS Aspergillus nidulans FGSC A4.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; Emericella.
OX NCBI_TaxID=227321;
RN [1]_NUCLEOTIDE SEQUENCE.
RC STRAIN=FGSC A4;
RA Birren B., Nusbaum C., Abouelleil A., Allen N., Anderson S.,
RA Arachchi H.M., Barna N., Bastien V., Bloom T., Boguslavsky L.,
RA Boukhgalter B., Butler J., Calvo S.E., Camarata J., Chang J.,
RA Choepel Y., Collymore A., Cook A., Cooke P., Corum B., DeArellano K.,
RA Diaz J.S., Dodge S., Dooley K., Dorris L., Elkins T., Engels R.,
RA Erickson J., Faro S., Ferreira P., FitzGerald M., Gage D., Galagan J.,
RA Gardyna S., Gnerre S., Graham L., Grand-Pierre N., Hafez N.,
RA Hagopian D., Hagos B., Hall J., Horton L., Hulme W., Iliev I.,
RA Jaffe D., Johnson R., Jones C., Kamal M., Kamat A., Karatas A.,
RA Kells C., Landers T., Levine R., Lindblad-Toh K., Liu G., Lui A.,
RA Ma L.-J., Mabbitt R., MacLean C., Macdonald P., Major J., Manning J.,
RA Matthews C., Mauceli E., McCarthy M., Meldrim J., Meneus L.,
RA Mihova T., Mlenga V., Murphy T., Naylor J., Nguyen C., Nicol R.,
RA Nielsen C.B., Norbu C., O'Connor T., O'Donnell P., O'Neil D.,
RA Oliver J., Peterson K., Phunkhang P., Pierre N., Purcell S.,
RA Rachupka A., Ramasamy U., Raymond C., Retta R., Rise C., Rogov P.,
RA Roman J., Schauer S., Schuback R., Seaman S., Severy P., Smirnov S.,
RA Smith C., Spencer B., Stange-Thomann N., Stojanovic N., Stubbs M.,
RA Talamas J., Tesfaye S., Theodore J., Topham K., Travers M.,
RA Vassiliev H., Venkataranan V.S., Viel R., Vo A., Wang S., Wilson B.,
RA Wu X., Wyman D., Young G., Zainoun J., Zembek L., Zimmer A., Zody M.,
RA Lander E.;
RT "Genomic Sequence of Aspergillus nidulans.";
RL Submitted (JAN-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; AACD01000172; EAA66464.1; -; Genomic_DNA.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
KW DNA-binding; Hypothetical protein; Nuclear protein.
SQ SEQUENCE 347 AA; 37841 MW; A7CAAFEE761B3B1 CRC64;

Query Match 51.6%; Score 176.5; DB 2; Length 347;
Best Local Similarity 57.4%; Pred. No. 4e-08;
Matches 39; Conservative 11; Mismatches 17; Indels 1; Gaps 1;

QY 1 KSTLPPRKAKTKEEQRRIILNRRAAHQSRKKELHLYLQYLRKCSLLENLNSVN 60
Db |:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|
71 KTNLPKRAKTEDEKEQRRIILNRRAAAQTSSRRKKLEKLEKIDMEQ-QNQFL 129

QY 61 LEKLADHE 68
Db |:|||||
130 LQRLAQME 137

RESULT 12
Q8TFUS_EMENI PRELIMINARY; PRT; 350 AA.
AC Q8TFUS;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Transcription factor.
GN Name=haa3;
OS Emericella nidulans (Aspergillus nidulans).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; Emericella.
OX NCBI_TaxID=162425;

Query Match 51.6%; Score 176.5; DB 2; Length 347;
Best Local Similarity 57.4%; Pred. No. 4e-08;
Matches 39; Conservative 11; Mismatches 17; Indels 1; Gaps 1;

QY 1 KSTLPPRKAKTKEEQRRIILNRRAAHQSRKKELHLYLQYLRKCSLLENLNSVN 60
Db |:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|
71 KTNLPKRAKTEDEKEQRRIILNRRAAAQTSSRRKKLEKLEKIDMEQ-QNQFL 129

QY 61 LEKLADHE 68
Db |:|||||
130 LQRLAQME 137

RESULT 13
Q4WEY8 ASPFU PRELIMINARY; PRT; 433 AA.
AC Q4WEY8;
DT 13-SEP-2005 (TrEMBLrel. 31, Created)
DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
DE BZIP transcription factor (HacA), putative.
GN ORFNames=Afu3g04070;
OS Aspergillus fumigatus Af293.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
OX NCBI_TaxID=330879;
RN [1]_NUCLEOTIDE SEQUENCE.
RC STRAIN=Af293;
RA Nierman W., Bain A., Anderson M.J., Wortman J., Kim H., Stanley J.,
RA Arroya J., Berriman M., Abe K., Archer D.B., Bermejo C., Bennett J.,
RA Bowyer P., Chen D., Collins M., Coulson R., Davies R., Dyer P.S.,
RA Farman M., Fedorova N., Fedorova N., Feldblyum T.V., Fischer R.,
RA Foster N., Fraser A., Garcia J.L., Garcia M.J., Goble A.,
RA Goldman G.H., Goni K., Griffith-Jones S., Gwilliam R., Haas B.,
RA Haas H., Harris D., Horiuchi H., Huang J., Humphrey S., Jimenez J.,
RA Keller N., Khouri H., Kitamoto K., Kobayashi T., Kulkarni R.,
RA Kumagai T., Lafton A., Latge J.-P., Li W., Lord A., Lu C.,
RA Majoros W.H., May G.S., Miller B.L., Mohamoud Y., Molina M., Monod M.,
RA Mouny I., Mulligan S., Murphy L., O'Neill S., Paulsen I.,
RA Penava M.A., Perlea M., Price C., Pritchard B.L., Quail M.A.,
RA Rabinowitz E., Rawlins N., Rajandream M.-A., Reichard U.,
RA Renauld H., Robson G.D., Rodriguez de Cordoba S., Rodriguez-Pena J.M.,
RA Ronning C.M., Rutter S., Salzberg S.L., Sanchez M.,
RA Sanchez-Ferrero J.C., Saunders D., Seeger K., Squares R., Squares S.,
RA Takeuchi M., Tekala P., Turner G., Vazquez de Aldana C.R., Weidman J.,
RA White O., Woodward J., Yu J.-H., Fraser C., Galagan J.E., Asai K.,
RA Machida M., Hall N., Barrell B., Denning D.W.;
RT "Genomic sequence of the pathogenic and allergenic filamentous fungus
Aspergillus fumigatus.";
RL Submitted (MAY-2005) to the EMBL/GenBank/DBJ databases.
CC -! CAUTION: The sequence shown here is derived from an
EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
preliminary data.
CC EMBL; AAHF01000010; EAL86689.1; -; Genomic_DNA.
DR
```

```

DR InterPro: IPR011700; bZIP 2.
DR InterPro: IPR004827; TF_bZIP.
DR Pfam: PF07716; bZIP_2; 1.
DR SMART: SM00338; BRLZ; 1.
DR PROSITE: PS00217; bZIP; 1.
DR PROSITE: PS00036; bZIP_BASIC; UNKNOWN_1.
KW DNA-binding; Nuclear protein.
SQ SEQUENCE 433 AA; 48908 MW; 3742DD0B43E05C74 CRC64;

Query Match      51.6%; Score 176.5; DB 2; Length 433;
Best Local Similarity 55.9%; Pred. No. 4.9e-08;
Matches 38; Conservative 12; Mismatches 17; Indels 1; Gaps 1;

QY 1 KSTLPPRKAKTKEKEQRRIERILNRRRAHQSRKRLHLYLERKCSLLENLNSVN 60
DB 1 KSTLPPRKAKTKEKEQRRIERILNRRRAHQSRKRLHLYLERKCSLLENLNSVN 60
QY 61 LEKLADHE 68
DB 123 LQRLSQME 130

RESULT 14
Q6FLY3_CANGA PRELIMINARY; PRT; 329 AA.
AC Q6FLY3;
DT 05-JUL-2004 (TREMBLrel. 27, Created)
DT 05-JUL-2004 (TREMBLrel. 27, Last sequence update)
DE Candida glabrata strain CBS138 chromosome K complete sequence.
GN Ordered locus names: CAGL0K12540g;
OS Candida glabrata (Yeast) (Torulopsis glabrata).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Mitosporic Saccharomycetales; Candida.
OX NCBI_TaxID=5478;
RN [1]_
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=ATCC 2001 / CBS 138;
RX PubMed=15229592; DOI=10.1038/nature02579;
RA Dujon B., Sherman D., Fischer G., Durrens P., Casaregola S.,
RA Lafontaine I., de Montigny J., Marc C., Neuveglise C., Talla E.,
RA Goffard N., Frangoul L., Aigle M., Anthonard V., Babour A., Barbe V.,
RA Barnay S., Blanchin S., Beckerich J.-M., Beyne E., Bleykaesten C.,
RA Boisarame A., Boyer J., Cattolico L., Confantolieri F., de Daruvar A.,
RA Despons L., Fabre E., Fairhead C., Ferry-Dumazet H., Groppi A.,
RA Hantraye F., Hennequin C., Jauniaux N., Joyet P., Kachouri R.,
RA Kerrest A., Koszul R., Lemaire M., Lesur I., Ma L., Muller H.,
RA Nicaud J.-M., Nikolski M., Oztas S., Ozier-Kalogeropoulos O.,
RA Pellenz S., Potier S., Richard G.-P., Straub M.-L., Suleau A.,
RA Swennen D., Tekala F., Wesolowski-Louvel M., Westhof E., Wirth B.,
RA Zenlou-Meyer M., Zivanovic Y., Bolotin-Fukuhara M., Thierry A.,
RA Bouchier C., Caudron B., Scarpelli C., Gaillardin C., Weissenbach J.,
RA Wincker P., Souciet J.-L.;
RT "Genome evolution in yeasts.";
RL Nature 430:35-44(2004).
DR EMBL; CR380957; CAG61724.1; -; Genomic_DNA.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro: IPR011616; bZIP 1.
DR InterPro: IPR004827; TF_bZIP.
DR Pfam: PF00170; bZIP_1; 1.
DR SMART: SM00338; BRLZ; 1.
DR PROSITE: PS00217; bZIP; 1.
DR PROSITE: PS00036; bZIP_BASIC; 1.
KW Complete proteome; DNA-binding; Nuclear protein.
SQ SEQUENCE 329 AA; 37162 MW; 469614BF3AF2282 CRC64;

Query Match      50.6%; Score 173; DB 2; Length 329;
Best Local Similarity 50.8%; Pred. No. 7.9e-08;
Matches 33; Conservative 19; Mismatches 9; Indels 4; Gaps 1;

QY 4 LPPRKAKTKEKEQRRIERILNRRRAHQSRKRLHLYLERKCS-----LLENLNSV 59

```

```

Db 29 MPPRKAKTKEKEIRKQIRILNRRRAAQSRDRKRNRYVANLEKCNMKVLDQLQSKI 88
QY 60 NLEKL 64
Db 89 DIKSM 93

RESULT 15
Q4HTT5_GIBZE PRELIMINARY; PRT; 429 AA.
AC Q4HTT5;
DT 13-SEP-2005 (TREMBLrel. 31, Created)
DT 13-SEP-2005 (TREMBLrel. 31, Last sequence update)
DT 13-SEP-2005 (TREMBLrel. 31, Last annotation update)
DE Hypothetical protein.
GN ORFNames=FGL1623.1;
OS Gibberella zeae PH-1.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Hypocreomycetidae; Hypocreales; Nectriaceae; Gibberella.
OX NCBI_TaxID=229533;
RN [1]_
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=PH-1;
RA Birren B., Nusbaum C., Abouelleil A., Allen N., Anderson S.,
RA Arachchi H.M., Barna N., Bastien V., Bloom T., Boguslavskiy L.,
RA Boukhgalter B., Butler J., Calvo S.E., Camarata J., Chang J.,
RA Choepel Y., Collymore A., Cook A., Cooke P., Corum B., Dearellano K.,
RA Diaz J.S., Dodge S., Dooley K., Dorris L., Elkins T., Engels R.,
RA Erickson J., Fairo S., Ferreira P., FitzGerald M., Gage D., Galagan J.,
RA Gardyna S., Gnerre S., Graham L., Grand-Pierre N., Hafez N.,
RA Hagopian D., Hagoos B., Hall J., Horton L., Hulme W., Iliev I.,
RA Jaffe D., Johnson R., Jones C., Kamal M., Kamat A., Karatas A.,
RA Kells C., Landers T., Levine R., Lindblad-Toh K., Liu G., Lui A.,
RA Ma L.-J., Mabbitt R., Maclean C., Macdonald P., Major J., Manning J.,
RA Matthews C., Mauceli E., McCarthy M., Meldrim J., Meneus L.,
RA Minova T., Mlenga V., Murphy T., Naylor J., Nguyen C., Nicol R.,
RA Nielsen C.B., Norbu C., O'Connor T., O'Donnell P., O'Neill D.,
RA Oliver J., Peterson K., Phunkhang P., Pierre N., Purcell S.,
RA Rachupka A., Ramasamy U., Raymond C., Retta R., Rise C., Rogov P.,
RA Roman J., Schauer S., Schupback R., Seaman S., Severy P., Smirnov S.,
RA Smith C., Spencer B., Strange-Thomann N., Stojanovic N., Stubbs M.,
RA Talamas J., Tesfaye S., Theodore J., Topham K., Travers M.,
RA Vasiliev H., Venkataraman V.S., Viel R., Vo A., Wang S., Wilson B.,
RA Wu X., Wyman D., Young G., Zainoun J., Zembek L., Zimmer A., Zody M.,
RA Lander E.;
RT "Fusarium graminearum genome sequence.";
RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; AACM01000497; EAA78735.1; -; Genomic_DNA.
KW Hypothetical protein.
SQ SEQUENCE 429 AA; 46630 MW; 04CCDE40BD392901 CRC64;

Query Match      45.9%; Score 157; DB 2; Length 429;
Best Local Similarity 47.9%; Pred. No. 2.9e-06;
Matches 34; Conservative 10; Mismatches 9; Indels 18; Gaps 1;

QY 1 KSTLPP-----RKRATKEEKEQRRIERILNRRRAHQSRKRLHLYLERKCSLLENLNSV 42
DB 100 KTNLPFRYGVDTLRHVVDSTDTLNRKRAKTBDEKEQRVERVLRNRRRAQSRKRQREV 159
QY 43 QYLERKCSLLE 53
Db 160 EALEKNGEELS 170

Search completed: November 23, 2005, 03:25:00
Job time : 30.2212 secs

```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 23, 2005, 03:07:01 ; Search time 4.7788 Seconds
(without alignments)
1369.117 Million cell updates/sec

Title: US-10-663-450-60
Perfect score: 342
Sequence: 1 KSTLPFRKAKTKEKEQRR.....CSLLENLLNSVNLEKLADHE 68

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues
Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 80:.*
1: PIR1:.*
2: PIR2:.*
3: PIR3:.*
4: PIR4:.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	342	100.0	238	2 S78571	transcription fact
2	124.5	36.4	168	2 T50922	bZIP protein HY5 l
3	110.5	32.3	322	2 T12093	TGACG-motif bindin
4	104.5	30.6	322	2 T08592	TGACG-motif bindin
5	104.5	30.6	326	2 T08591	TGACG-motif bindin
6	91.5	26.8	331	2 T26807	hypothetical prote
7	91.5	26.8	333	2 T26808	hypothetical prote
8	89.5	26.2	515	2 A42140	box B-binding fact
9	89.5	26.2	516	2 A44494	CAMP-responsive el
10	88.5	25.9	325	2 S23007	CAMP response elem
11	88.5	25.9	326	2 A40120	CAMP-responsive en
12	88.5	25.9	327	2 A35769	CAMP response elem
13	88.5	25.9	327	2 S20955	regulatory protei
14	88.5	25.9	327	2 S22298	CAMP response elem
15	88.5	25.9	328	2 A35663	CAMP response elem
16	88.5	25.9	341	2 S42699	hypothetical prote
17	88.5	25.9	341	2 B35769	CAMP response elem
18	88.5	25.9	341	2 S03343	CAMP response elem
19	88.5	25.9	412	2 D86203	hypothetical prote
20	87.5	25.6	120	2 JX0307	cyclic AMP-respons
21	87	25.4	486	2 JC4028	activating transcr
22	86.5	25.3	108	2 A49317	inducible CAMP ear
23	86.5	25.3	229	2 A37944	CAMP response elem
24	86.5	25.3	278	2 S36101	CAMP response elem
25	86.5	25.3	344	2 JCS601	CAMP response elem
26	85.5	25.0	378	2 T07154	bZIP DNA-binding p
27	85.5	25.0	505	1 S05380	transcriptional fac
28	85	24.9	176	2 B90087	hypothetical prote
29	85	24.9	269	2 A46490	TCR V beta CRE-mot

30	85	24.9	360	2 T03373	probable G-box bin
31	84.5	24.7	327	2 S22299	CAMP response elem
32	84.5	24.7	393	2 S16321	light-induced prot
33	84	24.6	271	2 S12560	transcription fact
34	84	24.6	295	4 S36174	RNA binding protei
35	84	24.6	313	2 A34785	DNA-binding protei
36	84	24.6	358	2 C42026	cyclic AMP respons
37	84	24.6	389	1 A39429	CAMP response elem
38	84	24.6	448	2 A42026	CAMP response elem
39	84	24.6	456	2 B42026	cyclic AMP respons
40	83.5	24.4	297	2 T51273	promoter-binding f
41	83.5	24.4	600	2 T00759	hypothetical prote
42	83	24.3	688	2 T32750	hypothetical prote
43	82.5	24.1	246	2 T12585	DC3 promoter-bindi
44	82	24.0	483	2 S12741	transcription fact
45	82	24.0	620	2 T49067	transcription fact

ALIGNMENTS

RESULT 1

S78571
transcription factor HAC1 - yeast (Saccharomyces cerevisiae)
N;Alternate names: protein YFL031w
C;Species: Saccharomyces cerevisiae
C;Date: 13-Sep-1998 #sequence revision 13-Sep-1998 #text_change 09-Jul-2004
C;Accession: S78571; S56223; S53578
R;Murakami, Y.
submitted to the Protein Sequence Database, January 1998
A;Reference number: S78570
A;Accession: S78571
A;Molecule type: DNA
A;Residues: 1-238 <MUR>
A;Cross-references: UNIPROT:P41546; UNIPARC:UPI0000157F61; EMBL:D50617; MIPS:YFL031w
A;Note: this is a revision to the sequence from reference S56186
R;Murakami, Y.; Naitou, M.; Hagiwara, H.; Shibata, T.; Ozawa, M.; Sasanuma, S.I.; Sasanuma, S.I.;
submitted to the EMBL Data Library, May 1995
A;Description: Analysis of the nucleotide sequence of chromosome VI from Saccharomyces cerevisiae
A;Reference number: S56186
A;Accession: S56223
A;Molecule type: DNA
A;Residues: 1-191, 'RWOTHSA1', <MUM>
A;Cross-references: UNIPARC:UPI0000179B60; EMBL:D50617; NID:g836685; PID:g836723; MIPS:YFL031w
A;Note: this sequence has been revised in reference S78570
A;Note: this was believed to be the complete sequence of YFL031w
R;Nojima, H.; Leem, S.H.; Araki, H.; Sakai, A.; Nakashima, N.; Kanaoka, Y.; Ono, Y.
Nucleic Acids Res. 22, 5279-5288, 1994
A;Title: Hac1: a novel yeast bZIP protein binding to the CRE motif is a multicopy suppressor of the bZIP1 mutation
A;Reference number: S53578; MUID:95116316; PMID:7616617
A;Accession: S53578
A;Molecule type: DNA
A;Residues: 1-142, 'RLCRPRVCRFRVGRDPFGMAECLRRKMYQSRRLPYPTI', 183-220, 'AVITMTRKLQ' <NOJ>
A;Cross-references: UNIPARC:UPI000012C042; GB:D26506; NID:g633122; PID:BAA05513.1; PID:G633122
C;Genetics:
A;Gene: SGD:HAC1; IRE2; ERN4
A;Cross-references: SGD:S0001863; MIPS:YFL031w
A;Map position: 6L
A;Introns: 221/1
C;Keywords: DNA binding; nucleus; transcription factor

Query Match 100.0%; Score 342; DB 2; Length 238;
Best Local Similarity 100.0%; Pred. No. 1.8e-24;
Matches 68; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	1	KSTLPFRKAKTKEKEQRRIRILNRRRAAQSREKKLHLQYLRKCSLLENLSVN	60
DB	23	KSTLPFRKAKTKEKEQRRIRILNRRRAAQSREKKLHLQYLRKCSLLENLSVN	82
QY	61	LEKLADHE	68
DB	83	LEKLADHE	90


```
A;Status: preliminary; translated from GB/EMBL/DBJ
A;Molecule type: DNA
A;Residues: 1-333 <MIL>
A;Cross-references: UNIPROT:Q9U210; UNIPARC:UPI000017A523; EMBL:AL032627; PIDN:CAB54382.
A;Experimental source: clone Y41C4A
C;Genetics:
A;Gene: CBSP:Y41C4A.4b
A;Introns: 24/3; 50/2; 81/3; 161/1; 230/1; 294/3
Query Match 26.8%; Score 91.5; DB 2; Length 333;
Best Local Similarity 38.5%; Pred. No. 0.24;
Matches 20; Conservative 15; Mismatches 16; Indels 1; Gaps 1;
QY 13 KEKEQRRIERILNRRAAHQSRKKRLHLQYLRKCSLLENLLNSVNLEKL 64
Db 268 EDESRRKQVRLDKNREAKKCRKKKYYKCLNRRVSVLEN-QNKALTBEL 318
RESULT 8
A42140
box B-binding factor-2 - fruit fly (Drosophila melanogaster)
C;Species: Drosophila melanogaster
C;Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 09-Jul-2004
C;Accession: A42140; S24542
R;Abel, T.; Bhatt, R.; Maniatis, T.
Genes Dev. 6, 466-480, 1992
A;Title: A Drosophila CREB/ATF transcriptional activator binds to both fat body- and liver-specific enhancers
A;Reference number: A42140; MUID:92192458; PMID:1532159
A;Accession: A42140
A;Molecule type: mRNA
A;Status: preliminary
A;Residues: 1-515 <ABE>
A;Cross-references: UNIPROT:P29747; UNIPARC:UPI00001283CE; EMBL:X64429; NID:g11063; PIDN:F;432-475/Domain: fos/jun DNA-binding domain homology <FUD>
A;Note: sequence extracted from NCBI backbone (NCBIP:88160)
C;Genetics:
A;Gene: FlyBase:Bbbf2
A;Cross-references: FlyBase:FBgn0004848
C;Superfamily: CAMP-responsive element-binding transcription activator; fos/jun DNA-binding domain homology <FUD>
F;432-475/Domain: fos/jun DNA-binding domain homology <FUD>
Query Match 26.2%; Score 89.5; DB 2; Length 515;
Best Local Similarity 37.5%; Pred. No. 0.53;
Matches 24; Conservative 14; Mismatches 23; Indels 3; Gaps 2;
QY 6 PRKRATK-EEKEQRRIERILNRRAAHQSRKKRLHLQYLRKCSLLENLLNSVNLE 62
Db 428 PQKLPLTKAEKSLKIRRKIKNKISAQESRRKKKEYMDQLERRVELVTENHDYKKRLE 487
QY 63 KLAD 66
Db 488 GLEE 491
RESULT 9
A44494
CAMP-responsive element-binding transcription activator CREB-A - fruit fly (Drosophila melanogaster)
C;Species: Drosophila melanogaster
C;Date: 10-Jun-1993 #sequence_revision 18-Nov-1994 #text_change 09-Jul-2004
C;Accession: A44494
R;Smolik, S.M.; Rose, R.E.; Goodman, R.H.
Mol. Cell. Biol. 12, 4123-4131, 1992
A;Title: A cyclic AMP-responsive element-binding transcriptional activator in Drosophila
A;Reference number: A44494; MUID:92375081; PMID:1508208
A;Accession: A44494
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-516 <SMO>
A;Cross-references: UNIPROT:P29747; UNIPARC:UPI0000078073; GB:M87038; NID:g157118; PIDN:F;432-475/Domain: fos/jun DNA-binding domain homology <FUD>
A;Note: sequence extracted from NCBI backbone (NCBIN:111718, NCBIP:111719)
C;Genetics:
A;Gene: FlyBase:Creba
A;Cross-references: FlyBase:FBgn0004396
C;Superfamily: CAMP-responsive element-binding transcription activator; fos/jun DNA-binding domain homology <FUD>
F;433-476/Domain: fos/jun DNA-binding domain homology <FUD>
Query Match 26.2%; Score 89.5; DB 2; Length 516;
Best Local Similarity 37.5%; Pred. No. 0.53;
Matches 24; Conservative 14; Mismatches 23; Indels 3; Gaps 2;
QY 6 PRKRATK-EEKEQRRIERILNRRAAHQSRKKRLHLQYLRKCSLLENLLNSVNLE 62
Db 429 PQKLPLTKAEKSLKIRRKIKNKISAQESRRKKKEYMDQLERRVELVTENHDYKKRLE 488
QY 63 KLAD 66
Db 489 GLEE 492
RESULT 10
S23007
CAMP response element-binding protein CREB.2 - bovine
C;Species: Bos primigenius taurus (cattle)
C;Date: 08-Jun-1994 #sequence_revision 10-Nov-1995 #text_change 09-Jul-2004
C;Accession: S23007
R;Williams, L.; Kettmann, R.; Chen, G.; Portetelle, D.; Burny, A.; Derse, D.
DNA Seq. 1, 415-417, 1991
A;Title: Nucleotide sequence of the bovine cyclic-AMP responsive DNA binding protein (CREB)
A;Reference number: S23007; MUID:92119333; PMID:1837490
A;Accession: S23007
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-325 <WIL>
A;Cross-references: UNIPROT:P27925; UNIPARC:UPI0000167BE3; EMBL:X57031; NID:g285; PIDN:C;Keywords: DNA binding; transcription regulation
C;Keywords: DNA binding; transcription regulation
F;259-302/Domain: fos/jun DNA-binding domain homology <FUD>
Query Match 25.9%; Score 88.5; DB 2; Length 325;
Best Local Similarity 33.9%; Pred. No. 0.44;
Matches 20; Conservative 16; Mismatches 22; Indels 1; Gaps 1;
QY 6 PRKRATK-EEKEQRRIERILNRRAAHQSRKKRLHLQYLRKCSLLENLLNSVNLEKL 64
Db 256 PALPTQPAEEAAKREVRMLKMKREAAECRRKKKEYVKCLNRRVAVLEN-QNKTLTBEL 313
RESULT 11
A40120
CAMP-responsive enhancer-binding protein CREB - human
C;Species: Homo sapiens (man)
C;Date: 27-Mar-1992 #sequence_revision 27-Mar-1992 #text_change 15-Mar-2004
C;Accession: A40120
R;Hoefler, J.P.; Meyer, T.E.; Yun, Y.; Jameson, J.L.; Habener, J.F.
Science 242, 1430-1433, 1988
A;Title: Cyclic AMP-responsive DNA-binding protein: structure based on a cloned placental
A;Reference number: A40120; MUID:89072714; PMID:2974179
A;Accession: A40120
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-326 <HOE>
A;Cross-references: UNIPARC:UPI000017A520; GB:M27691
F;260-303/Domain: fos/jun DNA-binding domain homology <FUD>
Query Match 25.9%; Score 88.5; DB 2; Length 326;
Best Local Similarity 33.9%; Pred. No. 0.44;
Matches 20; Conservative 16; Mismatches 22; Indels 1; Gaps 1;
QY 6 PRKRATK-EEKEQRRIERILNRRAAHQSRKKRLHLQYLRKCSLLENLLNSVNLEKL 64
Db 257 PALPTQPAEEAAKREVRMLKMKREAAECRRKKKEYVKCLNRRVAVLEN-QNKTLTBEL 314
RESULT 12
A35769
CAMP response element-binding protein A - human
C;Species: Homo sapiens (man)
```

C>Date: 05-Oct-1990 #sequence_revision 05-Oct-1990 #text_change 09-Jul-2004
A:Accession: A37340; A35769
R:Waeber, G.; Meyer, T.E.; Hoeffler, J.P.; Habener, J.F.
Trans. Assoc. Am. Physicians 103, 28-37, 1990
A:Title: Diversification of cyclic AMP-responsive enhancer binding proteins generated by A:Reference number: A37340; PMID:92087371; PMID:1966745
A:Accession: A37340
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-327 <WAE>
A:Cross-references: UNIPROT:PI6220; UNIPARC:UPI000002A717; GB:S72459; NID:g240428; PIDN:R:Berikowitz, L.A.; Gilman, M.Z.
Proc. Natl. Acad. Sci. U.S.A. 87, 5258-5262, 1990
A:Title: Two distinct forms of active transcription factor CREB (cAMP response element b A:Reference number: A35769; PMID:90319091; PMID:2142528
A:Accession: A35769
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-327 <BER>
A:Cross-references: UNIPARC:UPI000002A717; GB:M34356; NID:G181042; PIDN:AAA35716.1; PID: C:Keywords: alternative splicing; DNA binding; transcription regulation
F:261-304/Domain: fos/jun DNA-binding domain homology <FJD>

Query Match 25.9%; Score 88.5; DB 2; Length 327;
Best Local Similarity 33.9%; Pred. No. 0.44;
Matches 20; Conservative 16; Mismatches 22; Indels 1; Gaps 1;

QY 6 PRKRAKTEEKQRRRIERILNRRAAHQSRKKRLHLOYLERKCSLLENLLNSVNLKL 64
Db 258 PALPTQPAEEAARKREVRMLKMKREAAECRRKKKEYVKCLNRAVAVLEN-QNKTLEEL 315

RESULT 13
S20955
regulatory protein CREB - mouse
N:Alternate names: cAMP responsive element-binding protein
C:Species: Mus musculus (house mouse)
C>Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 09-Jul-2004
C:Accession: S20955
R:Ruppert, S.; Cole, T.J.; Boshart, M.; Schmid, E.; Schuetz, G.
EMBO J. 11, 1503-1512, 1992
A:Title: Multiple mRNA isoforms of the transcription activator protein CREB: generation A:Reference number: S20955; MUID:92224889; PMID:1532935
A:Accession: S20955
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-327 <RUP>
A:Cross-references: UNIPROT:Q01147; UNIPARC:UPI0000004066; EMBL:X67727
C:Genetics:
A:Introns: 38/3; 87/3; 121/2; 169/1; 230/1; 280/2
F:261-304/Domain: fos/jun DNA-binding domain homology <FJD>

Query Match 25.9%; Score 88.5; DB 2; Length 327;
Best Local Similarity 33.9%; Pred. No. 0.44;
Matches 20; Conservative 16; Mismatches 22; Indels 1; Gaps 1;

QY 6 PRKRAKTEEKQRRRIERILNRRAAHQSRKKRLHLOYLERKCSLLENLLNSVNLKL 64
Db 258 PALPTQPAEEAARKREVRMLKMKREAAECRRKKKEYVKCLNRAVAVLEN-QNKTLEEL 315

RESULT 14
S22298
cAMP response element-binding protein CREB-delta - human
C:Species: Homo sapiens (man)
C>Date: 19-Feb-1994 #sequence_revision 10-Nov-1995 #text_change 09-Jul-2004
C:Accession: S22298
R:Short, M.L.; Manohar, C.F.; Furtado, M.R.; Chadge, G.D.; Wolinsky, S.M.; Thimmapaya, B Nucleic Acids Res. 19, 4290, 1991
A:Title: Nucleotide and derived amino-acid sequences of the CRE-binding proteins from ra A:Reference number: S22298; MUID:91334144; PMID:1831258
A:Accession: S22298
A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: mRNA
A:Residues: 1-327 <SHO>
A:Cross-references: UNIPROT:PI6220; UNIPARC:UPI000016A784; EMBL:X60003; NID:g30493; PIDN: A:Note: the nucleotide sequence was submitted to the EMBL Data Library, May 1991
C:Keywords: DNA binding; transcription regulation
F:261-304/Domain: fos/jun DNA-binding domain homology <FJD>

Query Match 25.9%; Score 88.5; DB 2; Length 327;
Best Local Similarity 33.9%; Pred. No. 0.44;
Matches 20; Conservative 16; Mismatches 22; Indels 1; Gaps 1;

QY 6 PRKRAKTEEKQRRRIERILNRRAAHQSRKKRLHLOYLERKCSLLENLLNSVNLKL 64
Db 258 PALPTQPAEEAARKREVRMLKMKREAAECRRKKKEYVKCLNRAVAVLEN-QNKTLEEL 315

RESULT 15
A35663
cAMP response element-binding protein, hepatic - rat
C:Species: Rattus norvegicus (Norway rat)
C>Date: 28-Sep-1990 #sequence_revision 28-Sep-1990 #text_change 15-Mar-2004
C:Accession: A35663
R:Quinn, P.G.; Granner, D.K.
Mol. Cell. Biol. 10, 3357-3364, 1990
A:Title: Cyclic AMP-dependent protein kinase regulates transcription of the phosphoenolpy A:Reference number: A35663; MUID:90287125; PMID:2141384
A:Accession: A35663
A:Status: preliminary; nucleic acid sequence not shown; not compared with conceptual tra A:Molecule type: mRNA
A:Residues: 1-328 <QUI>
A:Cross-references: UNIPARC:UPI000017A525
C:Keywords: liver
F:262-305/Domain: fos/jun DNA-binding domain homology <FJD>

Query Match 25.9%; Score 88.5; DB 2; Length 328;
Best Local Similarity 33.9%; Pred. No. 0.44;
Matches 20; Conservative 16; Mismatches 22; Indels 1; Gaps 1;

QY 6 PRKRAKTEEKQRRRIERILNRRAAHQSRKKRLHLOYLERKCSLLENLLNSVNLKL 64
Db 259 PALPTQPAEEAARKREVRMLKMKREAAECRRKKKEYVKCLNRAVAVLEN-QNKTLEEL 316

Search completed: November 23, 2005, 03:26:04
Job time : 4.7788 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 22, 2005, 15:37:57 ; Search time 33.5 Seconds
(without alignments)
157.948 Million cell updates/sec

Title: US-10-663-450-6_COPY_53_116

Perfect score: 324

Sequence: 1 KPAKRRKSWGQELPVPKTN.....RAAAQTSRRERKLEMEKLES 64

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:*

- 1: /cgn2_6/prodata/1/iaa/5 COMB.pep.*
- 2: /cgn2_6/prodata/1/iaa/6 COMB.pep.*
- 3: /cgn2_6/prodata/1/iaa/H COMB.pep.*
- 4: /cgn2_6/prodata/1/iaa/PCTUS COMB.pep.*
- 5: /cgn2_6/prodata/1/iaa/RE COMB.pep.*
- 6: /cgn2_6/prodata/1/iaa/backfilese1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	277	85.5	409	2	US-09-533-029-104
2	166	51.2	248	2	US-09-248-796A-18840
3	100	30.9	143	2	US-09-640-211A-1930
4	91.5	28.2	102	2	US-09-640-211A-1046
5	91	28.1	264	2	US-09-949-016-10789
6	87.5	27.0	260	2	US-09-538-092-950
7	86	26.5	667	2	US-09-248-796A-14492
8	84.5	26.1	521	1	US-08-721-684C-2
9	84.5	26.1	521	1	US-09-005-970-2
10	84.5	26.1	521	2	US-09-407-715-2
11	83	25.6	23	2	US-09-831-642-72
12	80	24.7	212	2	US-09-489-039A-13074
13	79	24.4	63	2	US-09-107-433-3674
14	78	24.1	611	2	US-09-216-393B-81
15	77.5	23.9	1601	2	US-09-345-473B-40
16	77.5	23.9	1601	2	US-09-862-027-40
17	76.5	23.6	306	2	US-09-758-759-115
18	76.5	23.6	395	2	US-09-247-155-113
19	76.5	23.6	395	2	US-09-513-999C-14
20	76.5	23.6	395	2	US-09-471-276-14
21	76.5	23.6	395	2	US-09-903-190-113
22	75.5	23.3	351	2	US-09-870-089B-2
23	75.5	23.3	362	2	US-09-949-016-10923
24	75.5	23.3	501	2	US-09-949-016-11281
25	75.5	23.3	931	2	US-09-949-016-9850
26	74.5	23.0	338	1	US-08-218-686-2
27	74.5	23.0	338	2	US-08-460-242-2

28	74.5	23.0	452	2	US-09-949-016-7289
29	74	22.8	616	2	US-09-873-404-4
30	74	22.8	616	2	US-10-243-735-4
31	73	22.5	551	2	US-08-796-899-29
32	73	22.5	1637	2	US-09-718-692-2
33	73	22.5	1637	2	US-09-718-852-2
34	73	22.5	1637	2	US-09-718-815-2
35	72.5	22.4	371	2	US-09-148-545-259
36	72.5	22.4	371	2	US-09-621-011-259
37	72	22.2	645	2	US-09-949-016-11022
38	72	22.2	650	2	US-09-487-558B-430
39	72	22.2	700	2	US-09-831-642-34
40	71	21.9	325	2	US-09-267-031-14
41	71	21.9	802	2	US-09-823-240A-2
42	70.5	21.8	118	2	US-09-134-001C-2856
43	70.5	21.8	182	2	US-09-640-211A-800
44	70.5	21.8	307	2	US-09-267-031-6
45	70	21.6	472	2	US-09-520-781-22

ALIGNMENTS

RESULT 1

US-09-533-029-104
; Sequence 104, Application US/09533029
; Patent No. 6664446
; GENERAL INFORMATION:
; APPLICANT: Heard, Jacqueline
; APPLICANT: Broun, Pierre
; APPLICANT: Riechmann, Jose-Luis
; APPLICANT: Keddle, James
; APPLICANT: Pineda, Omaira
; APPLICANT: Adam, Luc
; APPLICANT: Samaha, Raymond
; APPLICANT: Zhang, James
; APPLICANT: Yu, Guo-liang
; APPLICANT: Ratcliffe, Oliver
; APPLICANT: Pilgrim, Marsha
; APPLICANT: Jiang, Cai-Zhong
; APPLICANT: Reuber, Lynne
; TITLE OF INVENTION: DISEASE-INDUCED POLYNUCLEOTIDES
; FILE REFERENCE: MBI-010
; CURRENT APPLICATION NUMBER: US/09/533,029
; EARLIER FILING DATE: 2000-03-22
; EARLIER APPLICATION NUMBER: 60/125,814
; NUMBER OF SEQ ID NOS: 121
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 104
; LENGTH: 409
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
; FEATURE:
; OTHER INFORMATION: G1034
US-09-533-029-104

Query Match 85.5%; Score 277; DB 2; Length 409;
Best Local Similarity 84.1%; Pred. No. 7.8e-25;
Matches 53; Conservative 5; Mismatches 5; Indels 0; Gaps 0;

QY	1	KPAKRRKSWGQELPVPKTNLPKRKRAKTEDEKEQRIERVLRNRAAAQTSRRERKLEME 60
DB	67	KKPIKKRKSQGQQLPEPTKTNLPKRKRAKTDQDEKEQRRVERVLRNRAAAQSSRRERKQVE 126
QY	61	KLE 63
DB	127	ALE 129

RESULT 2

US-09-248-796A-18840
; Sequence 18840, Application US/09248796A

```
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstock et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; FILE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.132
; CURRENT FILING DATE: 1999-02-12
; CURRENT APPLICATION NUMBER: US/09/248,796A
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; PRIOR FILING DATE: 1998-08-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 18840
; LENGTH: 248
; TYPE: PRT
; ORGANISM: Candida albicans
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: (221)
; OTHER INFORMATION: Identity of amino acid sequences at the above locations are unknown
US-09-248-796A-18840

Query Match      51.2%; Score 166; DB 2; Length 248;
Best Local Similarity 68.1%; Pred. No. 6.2e-12;
Matches 32; Conservative 8; Mismatches 7; Indels 0; Gaps 0;

QY 18 KTNLPKRKAKTEDEKQRIERVLNRRAAAQTSRERKLEMEKLES 64
Db 72 KSTLPKRKAKTQEEKQRIERVLNRRAAAHRSRERKQKHVEYLEN 118

RESULT 3
US-09-640-211A-1930
; Sequence 1930, Application US/09640211A
; Patent No. 6833446
; GENERAL INFORMATION:
; APPLICANT: Wood, Marion
; APPLICANT: Shenk, Michael A.
; APPLICANT: McGrath, Annette
; APPLICANT: Glenn, Matthew
; TITLE OF INVENTION: Compositions and Methods for the
; TITLE OF INVENTION: Modification of Gene Transcription
; FILE REFERENCE: 11000.1021C1U
; CURRENT APPLICATION NUMBER: US/09/640,211A
; CURRENT FILING DATE: 2000-08-16
; NUMBER OF SEQ ID NOS: 2368
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1930
; LENGTH: 143
; TYPE: PRT
; ORGANISM: Pinus radiata
US-09-640-211A-1930

Query Match      30.9%; Score 100; DB 2; Length 143;
Best Local Similarity 47.6%; Pred. No. 0.00022;
Matches 20; Conservative 8; Mismatches 14; Indels 0; Gaps 0;

QY 22 PPRKRAKTEDEKQRIERVLNRRAAAQTSRERKLEMEKLE 63
Db 35 PPRKARADLNAEQRREARHNRRAAQNRSRDKRKAQFTYME 76

RESULT 4
US-09-640-211A-1046
; Sequence 1046, Application US/09640211A
; Patent No. 6833446
; GENERAL INFORMATION:
; APPLICANT: Wood, Marion
; APPLICANT: Shenk, Michael A.
; APPLICANT: McGrath, Annette
; APPLICANT: Glenn, Matthew
; TITLE OF INVENTION: Compositions and Methods for the
```

```
; TITLE OF INVENTION: Modification of Gene Transcription
; FILE REFERENCE: 11000.1021C1U
; CURRENT APPLICATION NUMBER: US/09/640,211A
; CURRENT FILING DATE: 2000-08-16
; NUMBER OF SEQ ID NOS: 2368
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1046
; LENGTH: 102
; TYPE: PRT
; ORGANISM: Pinus radiata
US-09-640-211A-1046

Query Match      28.2%; Score 91.5; DB 2; Length 102;
Best Local Similarity 40.0%; Pred. No. 0.0015;
Matches 20; Conservative 11; Mismatches 18; Indels 1; Gaps 1;

QY 14 LPVPKTNLPKRKAKTEDEKQRIERVLNRRAAAQTSRERKLEMEKLE 63
Db 31 VPVP-AQANPKRPRSDLSAEKREARHNRRAAQNRSRDKRKAQFTSLE 79

RESULT 5
US-09-949-016-10789
; Sequence 10789, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; TITLE OF INVENTION: WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10789
; LENGTH: 264
; TYPE: PRT
; ORGANISM: Human
US-09-949-016-10789

Query Match      28.1%; Score 91; DB 2; Length 264;
Best Local Similarity 40.3%; Pred. No. 0.005;
Matches 27; Conservative 9; Mismatches 17; Indels 14; Gaps 4;

QY 11 GOELP--VP-----KTNLP-PRKRAK-TEDEKQRIERVLNRRAAAQTSRERKR 56
Db 36 GOALPLMVPQARGASPEAASGGLPQARKQRLTHLSPEKALRRKLNKRVAAQTARDKK 95

QY 57 LEMEKELE 63
Db 96 ARMSELE 102

RESULT 6
US-09-538-092-950
; Sequence 950, Application US/09538092
; Patent No. 6753314
; GENERAL INFORMATION:
; APPLICANT: Giot, Loic
; APPLICANT: Mansfield, Traci A.
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same
; FILE REFERENCE: 15966-542
; CURRENT APPLICATION NUMBER: US/09/538,092
; CURRENT FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: 60/127,352
; PRIOR FILING DATE: 1999-04-01
; PRIOR APPLICATION NUMBER: 60/178,965
```

```
; PRIOR FILING DATE: 2000-02-01
; NUMBER OF SEQ ID NOS: 1387
; SOFTWARE: CnraPatSeqFormat Version 0.9
; SEQ ID NO 950
; LENGTH: 260
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (0)...(0)
; OTHER INFORMATION: Polypeptide Accession Number P17861
US-09-538-092-950

Query Match      27.0%; Score 87.5; DB 2; Length 260;
Best Local Similarity 46.3%; Pred. No. 0.013;
Matches 19; Conservative 9; Mismatches 12; Indels 1; Gaps 1;

QY 24 RKRK-TEDEKORRIEVLNRRAAQTGRKRKLEMEKLE 63
Db 58 RKRQRLTHLSPKALURRLKRNVAQAQTARDRKARMSLE 98

RESULT 7
US-09-248-796A-14492
; Sequence 14492, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstein et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; CURRENT FILING DATE: 1999-02-12
; PRIOR FILING DATE: 1999-02-12
; PRIOR FILING DATE: 1998-02-13
; PRIOR APPLICATION NUMBER: US 60/074,725
; PRIOR FILING DATE: 1998-08-13
; PRIOR APPLICATION NUMBER: US 60/096,409
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 14492
; LENGTH: 667
; TYPE: PRT
; ORGANISM: Candida albicans
US-09-248-796A-14492

Query Match      26.5%; Score 86; DB 2; Length 667;
Best Local Similarity 33.3%; Pred. No. 0.056;
Matches 25; Conservative 13; Mismatches 25; Indels 12; Gaps 2;

QY 2 KPAKKGKSGQLPVPKTNLPK-----RAKTEDEKQORRI-----ERVLRNRAAAQ 49
Db 170 KPAKKGKAPAGLAALKQLKLEEQRRLEEEQRRLEEEERLAAEEAKKEAARA 229

QY 50 TSRERKRLEMEKLES 64
Db 230 AKKERLRKKEQLKA 244

RESULT 8
US-08-721-684C-2
; Sequence 2, Application US/08721684C
; Patent No. 5854016
; GENERAL INFORMATION:
; APPLICANT: Keegan, Kathleen S.
; TITLE OF INVENTION: No. 5854016el CREBa Isoform
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 233 South Wacker Drive, 6300 Sears Tower
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/005,970
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/721,684
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Williams Jr., Joseph A.
REGISTRATION NUMBER: 38,659
REFERENCE/DOCKET NUMBER: 27866/33487
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-474-6300
TELEFAX: 312-474-0448
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 521 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-005-970-2

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/721,684C
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Williams Jr., Joseph A.
; REGISTRATION NUMBER: 38,659
; REFERENCE/DOCKET NUMBER: 27866/33487
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-474-6300
; TELEFAX: 312-474-0448
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 521 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-721-684C-2

Query Match      26.1%; Score 84.5; DB 1; Length 521;
Best Local Similarity 35.6%; Pred. No. 0.064;
Matches 21; Conservative 13; Mismatches 20; Indels 5; Gaps 2;

QY 5 KKRKSGQLPVPKTNLPKRAKTEDEKQORRIEVLNRRAAQTGRKRKLEMEKLE 63
Db 270 EKRTLVAEGYPIP-TKLPLTK---SEBKALKIRKIKKIKSAQBSRRKKKEYMDSLE 323

RESULT 9
US-09-005-970-2
; Sequence 2, Application US/09005970
; Patent No. 5959079
; GENERAL INFORMATION:
; APPLICANT: Keegan, Kathleen S.
; TITLE OF INVENTION: No. 5959079el CREBa Isoform
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 233 South Wacker Drive, 6300 Sears Tower
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/005,970
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/721,684
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Williams Jr., Joseph A.
REGISTRATION NUMBER: 38,659
REFERENCE/DOCKET NUMBER: 27866/33487
TELECOMMUNICATION INFORMATION:
TELEPHONE: 312-474-6300
TELEFAX: 312-474-0448
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 521 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-005-970-2
```

```

Query Match      26.1%; Score 84.5; DB 1; Length 521;
Best Local Similarity 35.6%; Pred. No. 0.064;
Matches 21; Conservative 13; Mismatches 20; Indels 5; Gaps 2;

Qy 5 KKRKSWGQELPVPKTNLPPRKRAKTEDEKQRIERVLNRNAAQAOTSRERKRLEMEKLE 63
Db 270 EKRTLVAEGYPIP-TKLPLTK----SEKALKKIRRKIKNKISAQESRRKKKEYMDSLE 323

RESULT 10
US-09-407-715-2
; Sequence 2, Application US/09407715
; Patent No. 6248532
; GENERAL INFORMATION:
; APPLICANT: Keegan, Kathleen S.
; TITLE OF INVENTION: No. 6248532a1 CREBa Isoform
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
; STREET: 233 South Wacker Drive, 6300 Sears Tower
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/407,715
; FILING DATE: 28-Sep-1999
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/09/005,970
; FILING DATE: 12-JAN-1998
; APPLICATION NUMBER: US/08/721,684
; FILING DATE: 27-SEP-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Williams Jr., Joseph A.
; REGISTRATION NUMBER: 38,659
; REFERENCE/DOCKET NUMBER: 27866/33487
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312-474-6300
; TELEFAX: 312-474-0448
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 521 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 2:

US-09-407-715-2

Query Match      26.1%; Score 84.5; DB 2; Length 521;
Best Local Similarity 35.6%; Pred. No. 0.064;
Matches 21; Conservative 13; Mismatches 20; Indels 5; Gaps 2;

Qy 5 KKRKSWGQELPVPKTNLPPRKRAKTEDEKQRIERVLNRNAAQAOTSRERKRLEMEKLE 63
Db 270 EKRTLVAEGYPIP-TKLPLTK----SEKALKKIRRKIKNKISAQESRRKKKEYMDSLE 323

RESULT 11
US-09-831-642-72
; Sequence 72, Application US/09831642
; Patent No. 6635751
; GENERAL INFORMATION:
; APPLICANT: HAZE, Kyoosuke et al.
; TITLE OF INVENTION: ENDOPLASMIC RETICULUM STRESS TRANSCRIPTION FACTORS ATF6 AND CREB
; FILE REFERENCE: 1422-0474P
; CURRENT APPLICATION NUMBER: US/09/831,642
; CURRENT FILING DATE: 2001-05-11

```

```

; TITLE OF INVENTION: No. 6558903el Kinases and Uses Thereof
; FILE REFERENCE: 35800/183781
; CURRENT APPLICATION NUMBER: US/09/345,473E
; CURRENT FILING DATE: 1999-06-30
; NUMBER OF SEQ ID NOS: 62
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 40
; LENGTH: 1601
; TYPE: PRT
; ORGANISM: C. elegans
; US-09-345-473E-40

Query Match          23.9%; Score 77.5; DB 2; Length 1601;
Best Local Similarity 48.8%; Pred. No. 1.5;
Matches 20; Conservative 9; Mismatches 9; Indels 3; Gaps 2;

Oy      25  KRAKTEDE--KQQRIVRIVRNRAAAQTSRERKRLEMEKLE 63
        |||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db      461  KRAKBEERIREEAEITKEELRLRAAK-EKEKERLEKERLE 500

Search completed: November 22, 2005, 15:50:42
Job time : 33.5 secs

```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 22, 2005, 15:33:00 ; Search time 147.5 Seconds
(without alignments)
306.128 Million cell updates/sec

Title: US-10-663-450-6_COPY_53_116

Perfect score: 324

Sequence: 1 KKPAAKKRKSQGELPVPTKN.....RAAAQTSRERKRLMEKLES 64

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : UniProt 05.80.*

1: uniprot_sprot.*

2: uniprot_trembl.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	324	100.0	347	Q5AQN3_EMENI	Q5aqn3 aspergillus
2	324	100.0	350	Q8TFU8_EMENI	Q8tfu8 emericella
3	317	97.8	342	Q6W8X2_ASPNG	Q6w8x2 aspergillus
4	315	97.2	433	Q4WEY8_ASPFU	Q4wey8 aspergillus
5	283	87.3	451	Q8TFP3_TRIRE	Q8tfp3 trichoderma
6	282	87.0	430	Q7SHF0_NEUCR	Q7shf0 neurospora
7	265	81.8	556	Q51KW8_MAGGR	Q51kw8 magnaporthe
8	250	77.2	429	Q4HTY5_GIBZE	Q4hty5 gibberella
9	178.5	55.1	299	Q6CEV1_YARLI	Q6cevl yarrowia li
10	174.5	53.9	273	Q6CKQ1_KLULA	Q6ckq1 kluyveromyc
11	171.5	52.9	230	1 HAC1_YEAST	P41546 saccharomyc
12	171	52.8	260	Q6B0C2_DEBHA	Q6bgc2 debaryomyc
13	166	51.2	228	Q759Q5_ASHGO	Q75bq5 ashbya goss
14	166	51.2	357	Q5AA52_CANAL	Q5aa52 candida alb
15	144	44.4	329	Q6FLY3_CANGA	Q6fly3 candida gla
16	117	36.1	336	Q5VR11_ORYSA	Q5vr11 oryza sativ
17	116	35.8	168	1 HV5_ARATH	Q24646 arabidopsis
18	109	33.6	69	Q5H1L60_BRACM	Q5h1l60 brassica ca
19	109	33.6	109	Q84XX8_ERARP	Q84xx8 brassica ra
20	108	33.3	176	Q69XK6_ORYSA	Q69xk6 oryza sativ
21	105.5	32.6	141	Q677A7_9ASPA	Q677a7 hyacinthus
22	104.5	32.3	252	Q6NX18_XENTR	Q6nx18 xenopus tro
23	103.5	31.9	158	1 HV5_LYCES	Q9sm50 lycopersico
24	102.5	31.6	350	Q90Zr7_XENLA	Q90zr7 xenopus lae
25	102.5	31.6	396	Q7ZYC2_XENLA	Q7zyc2 xenopus lae
26	100	30.9	208	Q5DFK2_SCHJA	Q5dfk2 schistosoma
27	100	30.9	321	Q8GRY7_LOTJA	Q8gry7 lotus japon
28	99	30.6	188	Q6ZHT8_ORYSA	Q6zht8 oryza sativ
29	98	30.2	120	Q682B6_ARATH	Q682b6 arabidopsis
30	98	30.2	149	1 HVH_ARATH	Q8w191 arabidopsis
31	97	29.9	322	Q39896_SOYEN	Q39896 glycine max

32	97	29.9	326	2	Q39895_SOYEN	Q39895 glycine max
33	95	29.3	703	2	Q4WC74_ASPFU	Q4wc74 aspergillus
34	94.5	29.2	309	2	Q4H2M2_CIOIN	Q4h2m2 ciona intes
35	94	29.0	322	2	O04234_VICFA	O04234 vicia faba
36	94	29.0	627	2	Q5BD44_EMENI	Q5bd44 aspergillus
37	92	28.4	263	2	Q8UVQ5_BRARE	Q8uvq5 brachydanio
38	92	28.4	263	2	Q90X27_BRARE	Q90x27 brachydanio
39	92	28.4	383	2	Q8QHJ5_BRARE	Q8qhj5 brachydanio
40	92	28.4	383	2	Q90XD3_BRARE	Q90xd3 brachydanio
41	92	28.4	646	2	Q6AU90_ORYSA	Q6au90 oryza sativ
42	91.5	28.2	260	2	Q6EZA7_OREMO	Q6eza7 oreochromis
43	91.5	28.2	260	2	Q6EZA8_OREMO	Q6eza8 oreochromis
44	91	28.1	261	1	XBPL_HUMAN	P17861 homo sapien
45	90.5	27.9	176	2	Q98RX1_GUITH	Q98rx1 guillardia

ALIGNMENTS

RESULT 1
Q5AQN3_EMENI PRELIMINARY; PRT; 347 AA.
AC Q5AQN3;

DT 10-MAY-2005 (TrEMBLrel. 30, Created)

DT 10-MAY-2005 (TrEMBLrel. 30, Last sequence update)

DT 10-MAY-2005 (TrEMBLrel. 30, Last annotation update)

DE Hypothetical protein.

GN ORFNames=AN9397.2;

OS Aspergillus nidulans FGSC A4.

OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;

OC Eurotiales; Trichocomaceae; Emericella.

OX NCBI_TaxID=227321;

RN [1]

RP NUCLEOTIDE SEQUENCE.

RC STRAIN=FGSC A4;

RA Birren B., Nusbaum C., Abouelleil A., Allen N., Anderson S.,

Arachchi H.M., Barna N., Bastien V., Bloom T., Boguslavskiy L.,

Boukhgalter B., Butler J., Calvo S.E., Camarata J., Chang J.,

Choepel Y., Collymore A., Cooke P., Corum B., DeArelano K.,

Diaz J.S., Dodge S., Dooley K., Dorris L., Elkins T., Engels R.,

Erickson J., Faro S., Ferreira P., FitzGerald M., Gage D., Galagan J.,

Gardynia S., Gnarre S., Graham L., Grand-Pierre N., Hafez N.,

Hagopian D., Hagos B., Hall J., Horton L., Hulme W., Iliev I.,

Jaffe D., Johnson R., Jones C., Kamal M., Kamat A., Karatas A.,

Kells C., Landers T., Levine R., Lindblad-Toh K., Liu G., Lui A.,

Ma L.-J., Mabbitt R., Maclean C., Macdonald P., Major J., Manning J.,

Matthews C., Mauceli E., McCarthy M., Meldrim J., Meneus L.,

Mihova T., Mlenga V., Murphy T., Naylor J., Nguyen C., Nicol R.,

Nielsen C.B., Norbu C., O'Connor T., O'Donnell P., O'Neil D.,

Oliver J., Peterson K., Phunkhang P., Pierre N., Purcell S.,

Rachupka A., Ramasamy U., Raymond C., Retta R., Rise C., Rogov P.,

Roman J., Schauer S., Schupack R., Seaman S., Severy P., Smirnov S.,

Smith C., Spencer B., Stange-Thomann N., Stojanovic N., Stubbs M.,

Talanas J., Tesfaye S., Theodore J., Topham K., Travers M.,

Vassiliev H., Venkataraman V.S., Viel R., Vo A., Wang S., Wilson B.,

Wu X., Wyman D., Young G., Zainoun J., Zembek L., Zimmer A., Zody M.,

Lander E.;

RL "Genome Sequence of Aspergillus nidulans.";

RT Submitted (JAN-2004) to the EMBL/GenBank/DBJ databases.

DR EMBL; AACD01000172; EAA66464.1; -; Genomic_DNA.

DR GO; GO:0005634; C:nucleus; IEA.

DR GO; GO:0003677; F:DNA binding; IEA.

DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.

DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.

DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.

DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.

DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.

DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.

DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.

DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.

DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.

DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.

DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.

DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.

DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.

DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.

```
QY 61 KLES 64
    ||||
Db 114 KLES 117

RESULT 2
Q8TFUB EMENI
ID Q8TFUB_EMENI PRELIMINARY; PRT; 350 AA.
AC Q8TFUB;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Transcription factor.
GN Name=hacA;
OS Emericella nidulans (Aspergillus nidulane).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eutotiales; Trichocomaceae; Emericella.
OX NCBI_TaxID=162425;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Saloheimo M.L.A., Valkonen M., Penttilae M.E.;
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ413273; CAC88375.1; -; Genomic_DNA.
DR HSSP; P05412; 1JNM.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003677; F:DNA binding; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR011700; bZIP 2.
DR InterPro; IPR004827; TF_bZIP.
DR Pfam; PF07716; bZIP 2; 1.
DR SMART; SM00338; BRLZ; 1.
DR PROSITE; PS00217; bZIP; 1.
DR PROSITE; PS00036; bZIP_BASIC; UNKNOWN_1.
KW DNA-binding; Nuclear protein.
SQ SEQUENCE 350 AA; 38192 MW; 4F78340096064F44 CRC64;

Query Match 100.0%; Score 324; DB 2; Length 350;
Best Local Similarity 100.0%; Pred. No. 7.3e-22;
Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KKPAAKRRKSWGQELPVPKTNLPKRKRAKTEDEKEQRRIERVLRNRAAAQTSSRRKLEME 60
Db 54 KKPAAKRRKSWGQELPVPKTNLPKRKRAKTEDEKEQRRIERVLRNRAAAQTSSRRKLEME 113
QY 61 KLES 64
    ||||
Db 114 KLES 117

RESULT 3
Q6W8X2 ASPNG
ID Q6W8X2_ASPNG PRELIMINARY; PRT; 342 AA.
AC Q6W8X2;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Transcription factor HAC.
GN Name=hacA;
OS Aspergillus niger.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eutotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
OX NCBI_TaxID=5061;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RA Mulder H.J.;
RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY303684; AQA73495.1; -; Genomic_DNA.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0006377; F:DNA binding; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR011700; bZIP 2.
DR InterPro; IPR004827; TF_bZIP.

Query Match 97.8%; Score 317; DB 2; Length 342;
Best Local Similarity 96.9%; Pred. No. 3.1e-21;
Matches 62; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 KKPAAKRRKSWGQELPVPKTNLPKRKRAKTEDEKEQRRIERVLRNRAAAQTSSRRKLEME 60
Db 47 KKPAAKRRKSWGQELPVPKTNLPKRKRAKTEDEKEQRRIERVLRNRAAAQTSSRRKLEME 106
QY 61 KLES 64
    ||||
Db 107 KLEN 110

RESULT 4
Q4WEY8 ASPFU
ID Q4WEY8_ASPFU PRELIMINARY; PRT; 433 AA.
AC Q4WEY8;
DT 13-SEP-2005 (TrEMBLrel. 31, Created)
DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
DT 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
DE bZIP transcription factor (HacA), putative.
GN ORFNames=Afu3G04070;
OS Aspergillus fumigatus Af293.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eutotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
OX NCBI_TaxID=330879;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=Af293;
RA Nierman W., Pain A., Anderson M.J., Wortman J., Kim H.Stanley.,
RA Arroya J., Berriman M., Abe K., Archer D.B., Bermejo C., Bennett J.,
RA Bowyer P., Chen D., Collins M., Coulson R., Davies R., Dyer P.S.,
RA Farman M., Fedorova N., Fedorova N., Feldblyum T.V., Fischer R.,
RA Foster N., Fraser A., Garcia J.L., Garcia M.J., Gobie A.,
RA Goldman G.H., Goni K., Griffith-Jones S., Gwilliam R., Haas B.,
RA Haas H., Harris D., Horiuchi H., Huang J., Humphrey S., Jimenez J.,
RA Keller N., Khouri H., Kitamoto K., Kobayashi T., Kulkarni R.,
RA Kumagai T., Lafton A., Latge J.-P., Li W., Lord A., Lu C.,
RA Majoros W.H., May G.S., Miller B.B., Mohamoud Y., Molina M., Monod M.,
RA Mouyria I., Mulligan S., Murphy L., O'Neil S., Paulsen I.,
RA Penalba M.A., Perte M., Price C., Pritchard B.D., Quail M.A.,
RA Rabinowitsch E., Rawlins N., Rajandream M.-A., Reichard U.,
RA Renauld H., Robson G.D., Rodriguez de Cordoba S., Rodriguez-Pena J.M.,
RA Ronning C.M., Rutter S., Salzberg S.L., Sanchez M.,
RA Sanchez-Ferrero J.C., Saunders D., Seeger K., Squares R., Squares S.,
RA Takeuchi M., Tekala F., Turner G., Vazquez de Aldana C.R., Weidman J.,
RA White O., Woodward J., Yu J.-H., Fraser C., Galagan J.E., Asai K.,
RA Machida M., Hall N., Barrell B., Denning D.W.;
RT "Genomic sequence of the pathogenic and allergenic filamentous fungus
RT Aspergillus fumigatus.";
RL Submitted (MAY-2005) to the EMBL/GenBank/DBJ databases.
CC -!- CAUTION: the sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; AAHF01000010; EAL86689.1; -; Genomic_DNA.
DR InterPro; IPR011700; bZIP 2.
DR InterPro; IPR004827; TF_bZIP.
DR Pfam; PF07716; bZIP 2; 1.
DR SMART; SM00338; BRLZ; 1.
DR PROSITE; PS00217; bZIP; 1.
DR PROSITE; PS00036; bZIP_BASIC; UNKNOWN_1.
KW DNA-binding; Nuclear protein.
SQ SEQUENCE 433 AA; 48908 MW; 3742DD0B43E05C74 CRC64;

Query Match 97.2%; Score 315; DB 2; Length 433;
Best Local Similarity 95.3%; Pred. No. 6e-21;
```

RA Selitrennikoff C.P., Kinsey J.A., Braun E.L., Zelter A., Schulte U.,
RA Kothe G.O., Jedd G., Mewes W., Staben C., Marcotte E., Greenberg D.,
RA Roy A., Foley K., Naylor J., Thomann N., Barrett R., Gnerre S.,
RA Kamal M., Kamvysseisen M., Mauceli E., Bielek C., Ruck S., Frishman D.,
RA Kryštofova S., Rasmussen C., Metznerberg R.L., Perkins D.D., Kroken S.,
RA Cogoni C., Macino G., Catcheside D., Li W., Pratt R.J., Omani S.A.,
RA Desouza C.C., Glass L., Orbach M.J., Berglund J., Voelker R.,
RA Yarden O., Planann M., Seiler S., Dunlap J., Radford A., Aramayo R.,
RA Natvig D.O., Alex L.A., Mannhaupt G., Ebole D.J., Freitag M.,
RA Paulsen I., Sachs M.S., Lander E.S., Nussbaum C., Birren B.,
RT "The Genome Sequence of the Filamentous Fungus Neurospora crassa";
EL Nature 0-0-0(2003).
CC -1- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; AAEK01000004; C:nucleus; IEA.
DR GO; GO:0005634; C:nucleus; IEA.
DR GO; GO:0003677; P:DNA binding; IEA.
DR GO; GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro; IPR011700; bZIP_2.
DR InterPro; IPR004827; TF_bZIP.
DR Pfam; PF07716; bZIP_2; 1.
DR PROSITE; PSS0217; bZIP; 1.
DR SEQUENCE 430 AA; 46599 MW; 0EE0657CFA6160DA CRC64;
Query Match 87.0%; Score 282; DB 2; Length 430;
Best Local Similarity 87.3%; Pred. No. 6.5e-18;
Matches 55; Conservative 3; Mismatches 5; Indels 0; Gaps 0;
Qy 1 KKPAAKRSKSGQLPVPKTNLPKRKAQTEDEKQRIERVLNRRAAQTSSRRKRLEME 60
Db 88 KKPVRKRSKSGQLPVPEKTNLPKRKAQTEDEKQRRVERVLNRRAAQTSSRRKRLEVE 147
Qy 61 KLE 63
Db 148 GLE 150
RESULT 7
Q51KW8_MAGGR PRELIMINARY; PRT; 556 AA.
ID Q51KW8_MAGGR PRELIMINARY;
AC Q51KW8;
DT 13-SEP-2005 (TrEMBLrel. 31, Created)
DT 13-SEP-2005 (TrEMBLrel. 31, Last sequence update)
DE 13-SEP-2005 (TrEMBLrel. 31, Last annotation update)
DE Hypothetical protein.
GN ORFNames-MG9010.4;
OS Magnaporthe oryzae 70-15.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariomycetes incertae sedis; Magnaportheaceae; Magnaporthe.
OX NCBI_TaxID=242507;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=70-15;
RA Birren B., Nussbaum C., Abebe A., Abouelleil A., Adekoya E.,
RA Ait-zahra M., Allen T., Allen T., An P., Anderson M., Anderson S.,
RA Arachihi H., Armsbruster J., Bachantsang P., Baldwin J., Barry A.,
RA Bayul T., Blitshsteyn B., Bloom T., Blye J., Boguslavskiy L.,
RA Borowsky M., Boukgalter B., Brunache A., Butler J., Calixte N.,
RA Calvo S., Camarata J., Campo K., Chang J., Chesnatsang Y., Citroen M.,
RA Callimore A., Considine T., Cook A., Cooke P., Corum B., Cuomo C.,
RA David R., Dawoe T., Degray S., Dodge S., Dooley K., Dorje P.,
RA Dorjee K., Dorris L., Duffey N., Dupes A., Elkins T., Engle R.,
RA Erickson J., Farina A., Faro S., Ferreira P., Fischer H.,
RA Fitzgerald M., Foley K., Gage D., Galegan J., Gearin G., Gnerre S.,
RA Gnirke A., Goyette A., Graham J., Grandbois E., Gyaltsen K., Hafez N.,
RA Hagopian D., Hagos B., Hall J., Hatcher B., Heller A., Higgins H.,
RA Honan T., Horn A., Houde N., Hughes L., Hulme W., Husby E., Iliev I.,
RA Jaffe D., Jones C., Kamal M., Kamat A., Kamvysseisen M., Karlsson E.,
RA Kalls C., Kleu A., Kianer P., Kodira C., Kulbokas E., Labutti K.,
RA Lama D., Landers T., Leger J., Levine S., Lewis D., Lewis T.,
RA Lindblad-toh K., Liu X., Lokyitsang T., Lokyitsang Y., Lucien O.,
RA Lui A., Ma L.J., Mabbitt R., MacDonald J., Maclean C., Major J.,

RA Arachchi H.M., Barna N., Bastien V., Bloom T., Boguclavkiy L.,
 RA Boukhalter B., Butler J., Calvo S.E., Camarata J., Chang J.,
 RA Choepel Y., Collymore A., Cook A., Cooke P., Corum B., DeAtelliano K.,
 RA Diaz J.S., Dodge S., Dooley K., Dorris L., Elkins T., Engels R.,
 RA Erickson J., Faro S., Ferreira P., FitzGerald M., Gage D., Galagan J.,
 RA Gardyna S., Gnarr S., Graham L., Grand-Pierre N., Hafez N.,
 RA Hagopian D., Hagos B., Hall J., Horton L., Hulme W., Iliev I.,
 RA Jaffe D., Johnson R., Jones C., Kamal M., Kamat A., Karatas A.,
 RA Kells C., Lamberts T., Levine R., Lindblad-Toh K., Liu G., Lui A.,
 RA Ma L.-J., Mabbitt R., MacLean C., Macdonald P., Major J., Manning J.,
 RA Matthews C., Mauceli E., McCarthy M., Meldrim J., Meneus L.,
 RA Mihova T., Mengla V., Murphy T., Naylor J., Nguyen C., Nicol R.,
 RA Nielsen C.B., Norbu C., O'Connor T., O'Donnell P., O'Neill D.,
 RA Oliver J., Peterson K., Phunkhang P., Pierre N., Purcell S.,
 RA Rachupka A., Ramaeamy U., Raymond C., Retta R., Rise C., Rogov P.,
 RA Roman J., Schauer S., Schupback R., Seaman S., Severy P., Smirnov S.,
 RA Smith C., Spencer B., Stange-Thomann N., Stojanovic N., Stubbs M.,
 RA Talamas J., Tesfaye S., Theodore J., Topham K., Travers M.,
 RA Vassiliev H., Venkataraman V.S., Viel R., Vo A., Wang S., Wilson B.,
 RA Wu X., Wyman D., Young G., Zainoun J., Zembek L., Zimmer A., Zody M.,
 RA Lander E.;
 RT "Fusarium graminearum genome sequence";
 RL Submitted (FEB-2004) to the EMBL/GenBank/DBJ databases.
 CC -!- CAUTION: The sequence shown here is derived from an
 CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
 CC preliminary data.
 CC EMBL: AACM01000497; EAA78735.1; -; Genomic_DNA.
 CR Hypothetical protein.
 KW
 SQ SEQUENCE 429 AA; 46630 MW; 04CCDE40BD392901 CRC64;

 Query Match 77.2%; Score 250; DB 2; Length 429;
 Best Local Similarity 65.4%; Pred. No. 5.7e-15;
 Matches 53; Conservative 3; Mismatches 7; Indels 18; Gaps 1;

 QY 1 KKPAKRKKGQQLPVPKTNLP-----RYKATDEKEQRIRVRL 42
 Db 83 KKTTKRKKGQVLPPKTNLPYGDTLRHVVDESTDTLNRKRAKTEDEKEQRIRVRL 142

 QY 43 RNRRAAQTGRERKRLWEKLE 63
 Db 143 RNRRAAQSRRERKRQVEALE 163

 RESULT 9
 Q6CEV1_YARLI PRELIMINARY; PRT; 299 AA.
 AC Q6CEV1;
 DT 25-OCT-2004 (TrEMBLrel. 28, Created)
 DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
 DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
 DE Yarrowia lipolytica chromosome B of strain CLIB99 of Yarrowia
 DE lipolytica.
 GN OrderedLocusNames=YALI0B12716g;
 OS Yarrowia lipolytica (Candida lipolytica).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Dipodascaceae; Yarrowia.
 OX NCBI_TaxID=4952;
 RL [1]
 RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
 RC STRAIN=CLIB 122 / E 150;
 RC PubMed=15229592; DOI=10.1038/nature02579;
 RA Dujon B., Sherman D., Fischer J., Marck C., Neugeuglise C., Talla E.,
 RA Lafontaine I., de Montigny J., Marck C., Anthouard V., Babour A., Barbe V.,
 RA Goffard N., Frangeul L., Aigle M., Anthouard V., Babour A., Barbe V.,
 RA Barnay S., Blanchin S., Beckerich J.-M., Beyne E., Bleykascen C.,
 RA Boissame A., Boyer J., Cattolico L., Canfiolieri F., de Daruvar A.,
 RA Despons L., Fabre E., Fairhead C., Ferry-Dumazet H., Groppi A.,
 RA Denkaye F., Hennequin C., Jauniaux N., Joyet P., Kachouri R.,
 RA Kerrest A., Koszul R., Lemaire M., Lesur I., Ma L., Muller H.,
 RA Nicaud J.-M., Nikolski M., Ortiz S., Ozier-Kalagropoulos O.,
 RA Pellenn S., Potier S., Richard G.-F., Straub M.-L., Suleau A.,
 RA Swennen D., Tekala F., Wesolowski-Louvel M., Westhof E., Wirth B.,
 RA Zeniou-Meyer M., Zivanovic Y., Bolotin-Fukuhara M., Thierry A.,

Query Match	81.8%;	Score 265;	DB 2;	Length 556;
Best Local Similarity	82.5%;	Pred. NO. 3.1e-16;		
Matches 52;	Conservative 4;	Mismatches 7;	Indels 0;	Gaps 0;

QY	1	KPKAKRKSQWGQELPVPKTNLP	PPRKRAKTEDEKQRIERVLNR	AAQAQTSRRK	KELEWE 60
	:				
Db	88	EKKTKRKSQWGQVPEPKTNLP	PPRKRAKTEDEKQRRVERVLNR	RRAAQSRRK	QREVE 147
	:				
QY	61	KLE 63			
Db	148	ALE 150			

RESULT 8	
Q4HTT5_GIBZE	
ID	Q4HTT5_GIBZE PRELIMINARY; PRT; 429 AA.
AC	Q4HTT5;
DT	13-SEP-2005 (TREMBLrel. 31, Created)
DT	13-SEP-2005 (TREMBLrel. 31, Last sequence update)
DT	13-SEP-2005 (TREMBLrel. 31, Last annotation update)
DE	Hypothetical protein.
GN	ORFNames=FG11623.1;
OS	Gibberella zeae PH-1.
OC	Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC	Hypocreomycetidae; Hypocreales; Nectriaceae; Gibberella.
OX	NCBI_TaxID=229533;
RP	[1]
RN	NUCLEOTIDE SEQUENCE.
RP	STRAIN=PH-1;
RC	Birren B., Nussbaum C., Abouelleil A., Allen N., Anderson S.
RA	

[illegible]


```

DR GO: GO:0005634; C:nucleus; IEA.
DR GO: GO:0003677; F:DNA binding; IEA.
DR GO: GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro: IPR011700; bZIP_2.
DR Pfam: PF07716; bZIP_2; 1.
DR PROSITE: PS00036; bZIP_BASIC; 1.
KW Complete proteome.
SQ SEQUENCE 228 AA; 24914 MW; 605A60D78633F038 CRC64;

Query Match 51.2%; Score 166; DB 2; Length 228;
Best Local Similarity 71.7%; Pred. No. 1.6e-07;
Matches 33; Conservative 6; Mismatches 7; Indels 0; Gaps 0;

QY 18 KTNLPKPKRAKTEDEKQRIERVLRNRAAAQTSRKRKLEMEKLE 63
   |:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 14 KSTLPKPKRAKTEDEKQRIERVLRNRAAAQTSRKRKLEMEKLE 59

RESULT 14
Q5AA52 CANAL PRELIMINARY; PRT; 357 AA.
AC Q5AA52;
DT 10-MAY-2005 (TrEMBLrel. 30, Created)
DT 10-MAY-2005 (TrEMBLrel. 30, Last sequence update)
DT 10-MAY-2005 (TrEMBLrel. 30, Last annotation update)
DE Hypothetical protein HAC1.
GN Name=HAC1; ORFNames=CaO19.2432, CaO19.9968;
OS Candida albicans SC5314.
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX NCBI_TaxID=237561;
RN 1
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=SC5314;
RX PubMed=15123810; DOI=10.1073/pnas.0401648101;
RA Jones T., Federapfel N.A., Chibana H., Dungan J., Kalman S., Magee P.T.,
RA Magee B.B., Newport G., Thorntonsen Y.R., Agabian N.,
RA Davis R.W., Scherer S.;
RT "The diploid genome sequence of Candida albicans."
RL Proc. Natl. Acad. Sci. U.S.A. 101:7329-7334(2004).
RN 2
RP NUCLEOTIDE SEQUENCE.
RC STRAIN=SC5314;
RA Dungan J., Kuo A., Newport G., Lan C.-Y., Iijima C., Adegbola O.,
RA Roberts J., Persson K., Donnelly S., Favorito S., Tsung K.-W.,
RA Jones T., Scherer S., Agabian N.;
RT "Annotation of the Genome of Candida albicans."
RL Submitted (APR-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; AACQ01000041; EAK99529.1; -; Genomic DNA.
DR EMBL; AACQ01000040; EAK99617.1; -; Genomic DNA.
DR GO: GO:0005634; C:nucleus; IEA.
DR GO: GO:0003677; F:DNA binding; IEA.
DR GO: GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
KW DNA-binding, Hypothetical protein; Nuclear protein.
SQ SEQUENCE 357 AA; 40127 MW; 8E1C633A2B4ECDAB CRC64;

Query Match 51.2%; Score 166; DB 2; Length 357;
Best Local Similarity 68.1%; Pred. No. 2.5e-07;
Matches 32; Conservative 8; Mismatches 7; Indels 0; Gaps 0;

QY 18 KTNLPKPKRAKTEDEKQRIERVLRNRAAAQTSRKRKLEMEKLES 64
   |:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 54 KSTLPKPKRAKTEDEKQRIERVLRNRAAAQTSRKRKLEMEKLES 100

RESULT 15
Q6FLY3 CANGA PRELIMINARY; PRT; 329 AA.
AC Q6FLY3;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)

```

```

DE Candida glabrata strain CBS138 chromosome X complete sequence.
GN OrderedLocusNames=CAGL0K12540g;
OS Candida glabrata (Yeast) (Torulopsis glabrata).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX NCBI_TaxID=5478;
RN 1
RP NUCLEOTIDE SEQUENCE [LARGE SCALE GENOMIC DNA].
RC STRAIN=ATCC 2001 / CBS 138;
RX PubMed=15229592; DOI=10.1038/nature02579;
RA Lafontaine I., Sherman D., Fischer G., Durrans P., Casaregola S.,
RA Dujon B., Sherman D., Fischer G., March C., Neuvéglise C., Talla E.,
RA Geoffard N., Frangeul L., Aigle M., Anthouard V., Babour A., Barbe V.,
RA Barnay S., Blanchin S., Beckerich J.-M., Beyne E., Breykassen C.,
RA Boisrame A., Boyer J., Cattolico L., Confanioli F., de Daruvar A.,
RA Despons L., Fabre E., Fairhead C., Ferry-Dumazet H., Groppi A.,
RA Hantraye F., Hennequin C., Jauniaux N., Joyet P., Kachouri R.,
RA Kerrest A., Koszul R., Lemaire M., Lesur I., Ma L., Muller H.,
RA Nicaud J.-M., Nikolski M., Oztas S., Ozier-Kalogeropoulos O.,
RA Pellenz S., Potier S., Richard G.-F., Straub M.-L., Suleau A.,
RA Zenonen D., Tekala F., Wesolowski-Louvel M., Westhof E., Wirth B.,
RA Zeniou-Meyer M., Zivanovic Y., Bolotin-Fukuhara M., Thierry A.,
RA Bouchier C., Caudron B., Scarpelli C., Gaillardin C., Weissenbach J.,
RA Wincker P., Souciet J.-L.;
RT "Genome evolution in yeasts."
RL Nature 430:35-44(2004).
DR EMBL; CR380957; CAG61724.1; -; Genomic DNA.
DR GO: GO:0005634; C:nucleus; IEA.
DR GO: GO:0003677; F:DNA binding; IEA.
DR GO: GO:0006355; P:regulation of transcription, DNA-dependent; IEA.
DR InterPro: IPR011616; bZIP_1.
DR InterPro: IPR004827; TF_bZIP.
DR Pfam: PF00170; bZIP_1; 1.
DR SMART; SM00338; BRLZ; 1.
DR PROSITE; PS00217; bZIP; 1.
DR PROSITE; PS00036; bZIP_BASIC; 1.
KW Complete proteome; DNA-binding; Nuclear protein.
SQ SEQUENCE 329 AA; 37162 MW; 469614BF37AF2282 CRC64;

Query Match 44.4%; Score 144; DB 2; Length 329;
Best Local Similarity 63.0%; Pred. No. 2.5e-05;
Matches 29; Conservative 9; Mismatches 8; Indels 0; Gaps 0;

QY 18 KTNLPKPKRAKTEDEKQRIERVLRNRAAAQTSRKRKLEMEKLE 63
   |:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 26 ETWMPKPKRAKTEDEKQRIERVLRNRAAAQTSRKRKLEMEKLE 71

Search completed: November 22, 2005, 15:48:42
Job time : 148.5 secs

```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 22, 2005, 15:34:11 ; Search time 24 Seconds
(without alignments)
256.578 Million cell updates/sec

Title: US-10-663-450-6_COPY_53_116
Perfect score: 324
Sequence: 1 KIPAKKRSWGQELPVPKTN.....RAAAQTSRRKRLEMEKLES 64
Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues
Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR 80:.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	171.5	52.9	238	2 S78571	transcription fact
2	116	35.8	168	2 T09222	bZIP protein HY5 (
3	97	29.9	322	2 T08592	TGACG-motif-bindin
4	97	29.9	326	2 T08591	TGACG-motif bindin
5	94	29.0	322	2 T12093	TGACG-motif bindin
6	90.5	27.9	176	2 B90087	hypothetical prote
7	90.5	27.9	506	2 D84609	hypothetical prote
8	87.5	27.0	260	1 A36299	transcription fact
9	87.5	27.0	266	2 JC7300	tax-responsive ele
10	87.5	27.0	267	2 JC4857	hepatocarcinogenes
11	86	26.5	600	2 T00759	hypothetical prote
12	83	25.6	688	2 T32750	hypothetical prote
13	82.5	25.5	515	2 A42140	box B-binding fact
14	82.5	25.5	516	2 A44494	cAMP-responsive el
15	81	25.0	360	2 T03373	probable G-box bin
16	81	25.0	468	2 S33222	transcription fact
17	81	25.0	483	2 S12741	transcription fact
18	79.5	24.5	486	2 JC4028	activating transcr
19	77.5	23.9	1851	2 T19964	hypothetical prote
20	77	23.8	433	2 JC1230	DNA-binding protei
21	76.5	23.6	242	2 S05453	transcription fact
22	76	23.5	207	2 T40067	hypothetical prote
23	76	23.5	284	2 T24253	hypothetical prote
24	76	23.5	331	2 S33223	transcription fact
25	76	23.5	445	2 T50972	probable zotuin (i
26	75.5	23.3	246	2 T12585	Dc3 promoter-bindi
27	75.5	23.3	313	2 A34785	DNA-binding protei
28	75.5	23.3	349	2 A41349	histone-specific t
29	75.5	23.3	349	2 S77570	transcription fact

30	75.5	23.3	351	2 A45377	transcription fact
31	75.5	23.3	358	2 C42026	cyclic AMP respons
32	75.5	23.3	389	1 A39429	cAMP response elem
33	75.5	23.3	448	2 A42026	cAMP response elem
34	75.5	23.3	456	2 B42026	cyclic AMP respons
35	75.5	23.3	505	1 S05380	transcription fact
36	75.5	23.3	849	1 S64732	scaffold attachmen
37	75.5	23.3	1359	2 T34036	hypothetical prote
38	75	23.1	381	2 S26812	transcription fact
39	75	23.1	1549	1 A40691	trichohyalin - she
40	74.5	23.0	313	2 S66312	G-box binding fact
41	74.5	23.0	315	2 S20883	G-box-binding fact
42	74.5	23.0	315	2 G85433	G-box-binding fact
43	74.5	23.0	338	1 TVMSFB	transforming prote
44	74.5	23.0	338	2 IS3043	transforming prote
45	74.5	23.0	452	2 H96710	hypothetical prote

ALIGNMENTS

RESULT 1

S78571
transcription factor HAC1 - yeast (Saccharomyces cerevisiae)
N;Alternate names: protein YFL031w
C;Species: Saccharomyces cerevisiae
C;Date: 13-Sep-1998 #sequence_revision 13-Sep-1998 #text_change 09-Jul-2004
C;Accession: S78571; S56223; S53578
R;Murakami, Y.
submitted to the Protein Sequence Database, January 1998
A;Reference number: S78570
A;Accession: S78571
A;Molecule type: DNA
A;Residues: 1-238 <MUR>
A;Cross-references: UNIPROT:P41546; UNIPARC:UPI0000157F61; EMBL:D50617; MIPS:YFL031w
A;Note: this is a revision to the sequence from reference S56186
R;Murakami, Y.; Naitou, M.; Hagiwara, H.; Shibata, T.; Ozawa, M.; Sasanuma, S.I.; Sasanuma, S.I.;
submitted to the EMBL Data Library, May 1995
A;Description: Analysis of the nucleotide sequence of chromosome VI from Saccharomyces ce;
A;Reference number: S56186
A;Accession: S56223
A;Molecule type: DNA
A;Residues: 1-131, 'RWQTHSATI' <MUR>
A;Cross-references: UNIPARC:UPI0000179B60; EMBL:D50617; NID:g836685; PID:g836723; MIPS:YI
A;Note: this sequence has been revised in reference S78570
A;Note: this was believed to be the complete sequence of YFL031w
R;Nojima, H.; Leem, S.H.; Araki, H.; Sakai, A.; Nakashima, N.; Kanaoka, Y.; Ono, Y.
Nucleic Acids Res. 22, 5279-5288, 1994
A;Title: Hac1: a novel yeast bZIP protein binding to the CRE motif is a multicopy suppress;
A;Reference number: S53578; MUID:95116316; PMID:7816617
A;Accession: S53578
A;Molecule type: DNA
A;Residues: 1-142, 'RLCRPRYCRFRVGRDPFMAECURRMYOSRRRLPVTI', 183-220, 'AVITMTRKLO' <NOJ>
A;Cross-references: UNIPARC:UPI000012C042; GB:D26506; NID:g633122; PIDN:BAA05513.1; PID:;
C;Genetics:
A;Gene: SGD:HAC1; IRE2; ERN4
A;Cross-references: SGD:S0001863; MIPS:YFL031w
A;Map position: 6L
A;Introns: 221/1
C;Keywords: DNA binding; nucleus; transcription factor

Query Match 52.9%; Score 171.5; DB 2; Length 238;
Best Local Similarity 66.0%; Pred. No. 3.7e-09;
Matches 35; Conservative 8; Mismatches 7; Indels 3; Gaps 1;

QY 14 LPVP---KTNLPKRKTEDEKQRIEVLNRRAAAQTSRRKRLEMEKLE 63

DB 16 LAIPNFKSTLPPRKAKTKESKQRIEVLNRRAAAQTSRRKRLEMEKLE 68

RESULT 2

T50922
bZIP protein HY5 [imported] - Arabidopsis thaliana

A;Reference number: A84420; MUID:20083487; PMID:10617197

A;Accession: D84609

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-506 <STO>

A;Cross-references: UNIPROT:Q9SIF0; UNIPARC:UPI00000A51B3; GB:AE002093; PMID:94567209; PMID:94567209

C;Genetics:

A;Gene: At2g22140

A;Map position: 2

Query Match 27.9%; Score 90.5; DB 2; Length 506;

Best Local Similarity 34.8%; Pred. No. 0.29;

Matches 23; Conservative 13; Mismatches 25; Indels 5; Gaps 1;

QY 1 KKPAKKKSGQLPVKTLPRKRAKTE-----DEKQRRRIERVLRNRAAAQTSSRERK 55

Db 161 EKTGRKKKIRTTTLTPVGEALPKQSKEDKTSAMEBKLRKQERLEKAASAEERK 220

QY 56 RLEMEK 61

Db 221 RLSKEK 226

RESULT 8

A36299

transcription factor hXBP-1 - human

N;Alternate names: DNA-binding protein TREBS

C;Species: Homo sapiens (man)

C;Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 09-Jul-2004

C;Accession: A36299; S12559; A47490

R;Liou, H.C.; Boothby, M.R.; Finn, P.W.; Davidson, R.; Nabavi, N.; Zeleznik-Le, N.J.; Tinsley, M.R.;

Science 247, 1581-1583, 1990

A;Title: A new member of the leucine zipper class of proteins that binds to the HLA DRα1

A;Reference number: A36299; MUID:90208323; PMID:2321018

A;Accession: A36299

A;Status: preliminary

A;Molecule type: mRNA

A;Residues: 1-260 <LIO>

A;Cross-references: UNIPROT:P17861; UNIPARC:UPI0000031C60; GB:M31627; NID:g184485; PIDN:EMBO J. 9, 2537-2542, 1990

R;Yoshimura, T.; Fujisawa, J.I.; Yoshida, M.

A;Title: Multiple cDNA clones encoding nuclear proteins that bind to the tax-dependent element

A;Reference number: S12559; MUID:90316112; PMID:2196176

A;Accession: S12559

A;Molecule type: DNA

A;Residues: 1-32, 'GQA', 35-260 <YOS>

A;Cross-references: UNIPARC:UPI000006FC05; EMBL:X55543; NID:g287644; PIDN:CAA39149.1; PMID:10617197

R;Ponath, P.D.; Fass, D.; Liou, H.C.; Glimcher, L.H.; Strominger, J.L.

J. Biol. Chem. 268, 17074-17082, 1993

A;Title: The regulatory gene, hXBP-1, and its target, HLA-DRA, utilize both common and distinct

A;Reference number: A47490; MUID:93352484; PMID:8349596

A;Accession: A47490

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-32, 'GQA', 35-128, 'T', 130-192, 'P', 194-198, 'R', 200-260 <PON>

A;Cross-references: UNIPARC:UPI000017330E; GB:L13850

A;Note: authors translated the codon ACC for residue 130 as Asn, AAG for residue 151 as

C;Genetics:

A;Gene: GDB:XBP1; XBP2

A;Cross-references: GDB:131393; OMIM:194355

A;Map position: 22pter-22qter

A;Introns: 76/3; 108/3; 151/3; 200/3

C;Superfamily: X box-binding protein 1; fos/jun DNA-binding domain homology

C;Keywords: DNA binding; leucine zipper; nucleus; transcription regulation

F;64-104/Domain: fos/jun DNA-binding domain homology <FJD>

Query Match

Best Local Similarity

Matches

QY

Db

27.0%; Score 87.5; DB 1; Length 260;

46.3%; Pred. No. 0.3;

19; Conservative

9; Mismatches

12; Indels

1; Gaps

1;

24 RKRAK-TEDEKQRRRIERVLRNRAAAQTSSRERKLEMEKLE 63

58 RKQRQLTHLSPEKALRRKLRKNRAAQTARDRKARMESE 98

RESULT 9

JC7300

tax-responsive element-binding protein 5 - mouse

C;Species: Mus musculus (house mouse)

C;Date: 18-Aug-2000 #sequence_revision 18-Aug-2000 #text_change 09-Jul-2004

C;Accession: JC7300

R;Masaki, T.; Noguchi, H.; Kobayashi, M.; Yoshida, M.; Takamatsu, K.

DNA Res. 7, 187-193, 2000

A;Title: Isolation and characterization of the gene encoding mouse tax-responsive element

A;Reference number: JC7300

A;Accession: JC7300

A;Molecule type: mRNA

A;Residues: 1-266 <NAS>

A;Cross-references: UNIPROT:Q9ESS3; UNIPARC:UPI00000E73B0; DDBJ:AB036745

C;Genetics:

A;Gene: trebs

C;Superfamily: X box-binding protein 1; fos/jun DNA-binding domain homology

C;Keywords: leucine zipper; transcription factor

Query Match

Best Local Similarity

Matches

QY

Db

27.0%; Score 87.5; DB 2; Length 266;

46.3%; Pred. No. 0.31;

19; Conservative

9; Mismatches

12; Indels

1; Gaps

1;

24 RKRAK-TEDEKQRRRIERVLRNRAAAQTSSRERKLEMEKLE 63

51 RKQRQLTHLSPEKALRRKLRKNRAAQTARDRKARMESE 91

RESULT 10

JC4857

hepatocarcinogenesis-related transcription factor - rat

C;Species: Rattus norvegicus (Norway rat)

C;Date: 15-Aug-1996 #sequence_revision 18-Oct-1996 #text_change 09-Jul-2004

C;Accession: JC4857

R;Kishimoto, T.; Kokura, K.; Kumagai, Y.; Makino, Y.; Tamura, T.

Biochem. Biophys. Res. Commun. 224, 746-751, 1996

A;Title: HTF: A b-zip transcription factor that is closely related to the human XBP/TREB

A;Reference number: JC4857

A;Accession: JC4857

A;Molecule type: mRNA

A;Residues: 1-267 <KIS>

A;Cross-references: UNIPROT:Q9R1S4; UNIPARC:UPI00000B80AD

C;Comment: This is a basic-leucine zipper type transcription factor involved in hepatocel

C;Genetics:

A;Gene: htf

C;Superfamily: X box-binding protein 1; fos/jun DNA-binding domain homology

C;Keywords: leucine zipper; transcription factor

F;58-98/Domain: fos/jun DNA-binding domain homology <FJD>

F;89-126/Region: leucine zipper motif

Query Match

Best Local Similarity

Matches

QY

Db

27.0%; Score 87.5; DB 2; Length 267;

46.3%; Pred. No. 0.31;

19; Conservative

9; Mismatches

12; Indels

1; Gaps

1;

24 RKRAK-TEDEKQRRRIERVLRNRAAAQTSSRERKLEMEKLE 63

52 RKQRQLTHLSPEKALRRKLRKNRAAQTARDRKARMESE 92

RESULT 11

T00759

hypothetical protein At2g40950 [imported] - Arabidopsis thaliana

N;Alternate names: hypothetical protein T20B5.15

C;Species: Arabidopsis thaliana (mouse-ear cress)

C;Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 31-Dec-2004

C;Accession: T00759; H84835

R;Rounsley, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; Kaul,

submitted to the EMBL Data Library, November 1997

A;Description: Arabidopsis thaliana chromosome II BAC T20B5 genomic sequence.

A;Reference number: Z14159

A;Accession: T00759

A:Status: translated from GB/EMBL/DBBJ
A:Molecule type: DNA
A:Residues: 1-600 <R0U>
A:Cross-references: UNIPROT:O22208; UNIPARC:UPI0000179D30; EMBL:AC002409; NID:g2623294;
A:Experimental source: cultivar Columbia
R:Lin, X.; Kaul, S.; Roundley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Mofat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.;
eugs, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
Nature 402, 761-768, 1999
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487; PMID:10617197
A:Accession: H84835
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-600 <STO>
A:Cross-references: UNIPARC:UPI0000179D30; GB:AE002093; NID:g2623309; PIDN:AAB86455.1; C
C:Genetics:
A:Gene: T20B5.15; At2g40950
A:Map position: 2
C:Superfamily: bZIP transcription factor; fos/jun DNA-binding domain homology

Query Match 26.5%; Score 86; DB 2; Length 600;
Best Local Similarity 30.8%; Pred. No. 0.91;
Matches 20; Conservative 18; Mismatches 21; Indels 6; Gaps 1;

QY 5 KKRKSGQELPVPTNLPPRKRAKTED-----EKEQRRIERVLNRNRAAAQTSRRKRKLE 58
DB 193 KRKKEIDELTDESRNSKYRRSGEDADASAVTGEDEKKRRLMRNESAQLSRQRKKHY 252

QY 59 MEKLE 63
DB 253 VEELE 257

RESULT 12
T32750
hypothetical protein F57B10.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 09-Jul-2004
C:Accession: T32750
R:Greco, T.; Elliott, G.; Keppler, D.
submitted to the EMBL Data Library, December 1997
A:Description: The sequence of C. elegans cosmid F57B10.
A:Reference number: Z21219
A:Accession: T32750
A:Status: preliminary; translated from GB/EMBL/DBBJ
A:Molecule type: DNA
A:Residues: 1-688 <GRE>
A:Cross-references: UNIPROT:O44743; UNIPARC:UPI0000079D06; EMBL:AF039713; PIDN:AAB96719.
A:Experimental source: strain Bristol N2; clone F57B10
C:Genetics:
A:Gene: CESP.F57B10.1
A:Map position: 1
A:Introns: 54/1; 150/3; 196/1; 306/3; 383/3; 627/3

Query Match 25.6%; Score 83; DB 2; Length 688;
Best Local Similarity 48.5%; Pred. No. 2;
Matches 16; Conservative 11; Mismatches 6; Indels 0; Gaps 0;

QY 31 DEKEQRRIERVLNRNRAAAQTSRRKRKLEMEKLE 63
DB 281 EERDLKIRRKIRNKESAQTSRRKKQDYIEQLE 313

RESULT 13
A42140
box B-binding factor-2 - fruit fly (Drosophila melanogaster)
C:Species: Drosophila melanogaster
C:Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 09-Jul-2004
C:Accession: A42140; S24542
R:Abel, T.; Bhatt, R.; Maniatis, T.
Genes Dev. 6, 466-480, 1992
A:Title: A Drosophila CREB/ATF transcriptional activator binds to both fat body- and liv

A:Reference number: A42140; MUID:92192458; PMID:1532159
A:Accession: A42140
A:Status: Preliminary
A:Molecule type: mRNA
A:Residues: 1-515 <ABE>
A:Cross-references: UNIPROT:P29747; UNIPARC:UPI00001283CE; EMBL:X64429; NID:g11063; PIDN:
A:Note: sequence extracted from NCBI backbone (NCBIP:88160)
C:Genetics:
A:Gene: FlyBase:Bbbf2
A:Cross-references: FlyBase:FBgn0004848
C:Superfamily: cAMP-responsive element-binding transcription activator; fos/jun DNA-binding
C:Keywords: DNA binding; nucleus; transcription regulation
F:432-475/Domain: fos/jun DNA-binding domain homology <FJD>

Query Match 25.5%; Score 82.5; DB 2; Length 515;
Best Local Similarity 33.9%; Pred. No. 1.7;
Matches 20; Conservative 15; Mismatches 19; Indels 5; Gaps 2;

QY 5 KKRKSGQELPVPTNLPPRKRAKTEDEKEQRRIERVLNRNRAAAQTSRRKRKLEMEKLE 63
DB 416 EKRTLLAEGYPIQ-KLPLTK---AEEKSLKKIRKIKNKISAQESRRKKKEYMDQLE 469

RESULT 14
A44494
cAMP-responsive element-binding transcription activator CREB-A - fruit fly (Drosophila me
C:Species: Drosophila melanogaster
C:Date: 10-Jun-1993 #sequence_revision 18-Nov-1994 #text_change 09-Jul-2004
C:Accession: A44494
R:Smolik, S.M.; Rose, R.E.; Goodman, R.H.
Mol. Cell. Biol. 12, 4123-4131, 1992
A:Title: A cyclic AMP-responsive element-binding transcriptional activator in Drosophila
A:Reference number: A44494; MUID:92375081; PMID:1508208
A:Accession: A44494
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-516 <SMO>
A:Cross-references: UNIPROT:P29747; UNIPARC:UPI0000078073; GB:M87038; NID:g157118; PID:g
A:Note: sequence extracted from NCBI backbone (NCBIN:111718, NCBIP:111719)
C:Genetics:
A:Gene: FlyBase:CrebA
A:Cross-references: FlyBase:FBgn0004396
C:Superfamily: cAMP-responsive element-binding transcription activator; fos/jun DNA-binding
F:433-476/Domain: fos/jun DNA-binding domain homology <FJD>

Query Match 25.5%; Score 82.5; DB 2; Length 516;
Best Local Similarity 33.9%; Pred. No. 1.7;
Matches 20; Conservative 15; Mismatches 19; Indels 5; Gaps 2;

QY 5 KKRKSGQELPVPTNLPPRKRAKTEDEKEQRRIERVLNRNRAAAQTSRRKRKLEMEKLE 63
DB 417 EKRTLLAEGYPIQ-KLPLTK---AEEKSLKKIRKIKNKISAQESRRKKKEYMDQLE 470

RESULT 15
T03373
probable G-box binding factor 8 - rice
C:Species: Oryza sativa (rice)
C:Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 31-Dec-2004
C:Accession: T03373
R:Nakagawa, H.; Ohmura, K.; Hattori, T.
Plant J. 9, 217-227, 1996
A:Title: A rice bZIP protein, designated OSB28, is rapidly induced by abscisic acid.
A:Reference number: Z14906; MUID:96417817; PMID:8820608
A:Accession: T03373
A:Status: preliminary; translated from GB/EMBL/DBBJ
A:Molecule type: mRNA
A:Residues: 1-360 <NAK>
A:Cross-references: UNIPROT:O40645; UNIPARC:UPI00000A8257; EMBL:U42208; NID:g1147631; PI
A:Experimental source: cv. Nipponbare
C:Superfamily: bZIP G-box-binding protein; fos/jun DNA-binding domain homology

Query Match 25.0%; Score 81; DB 2; Length 360;

Best Local Similarity 36.1%; Pred. No. 1.7;
 Matches 22; Conservative 10; Mismatches 23; Indels 6; Gaps 1;
 QY 2 KPAKKKSGQELVPKTNLPKRKRAKTEDEKQRIERVLRNRAAAQTSRERKRLMEK 61
 Db 199 KPDVSTASDFRVIATPVTEVP-----TKDDKSKRERKQSNRESARRSLRKAETEE 252
 QY 62 L 62
 Db 253 L 253

Search completed: November 22, 2005, 15:49:34
 Job time : 29 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: November 22, 2005, 15:31:54 ; Search time 159 Seconds
(without alignments)
176.857 Million cell updates/sec

Title: US-10-663-450-6_COPY_53_116
Perfect score: 324
Sequence: 1 KPAKRRKSGWQLPVPKTN.....RAAAQTSRRKRLEMEKLES 64

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_21.*
1: Geneseqp1980s.*
2: Geneseqp1990s.*
3: Geneseqp2000s.*
4: Geneseqp2001s.*
5: Geneseqp2002s.*
6: Geneseqp2003as.*
7: Geneseqp2003bs.*
8: Geneseqp2004s.*
9: Geneseqp2005s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	324	100.0	349	8 ADS12805	Ads12805 Aspergill
2	324	100.0	349	8 ADS12803	Ads12803 Aspergill
3	324	100.0	349	8 ADS12863	Ads12863 Aspergill
4	324	100.0	350	4 AAB82976	Aab82976 Aspergill
5	324	100.0	350	5 AAE15372	Aae15372 Aspergill
6	320	98.8	64	5 AAE15374	Aae15374 Aspergill
7	317	97.8	342	4 AAB82977	Aab82977 Aspergill
8	317	97.8	342	5 AAE15381	Aae15381 Aspergill
9	317	97.8	342	8 ADS12818	Ads12818 Aspergill
10	317	97.8	386	5 AAE15379	Aae15379 Aspergill
11	317	97.8	386	8 ADS12815	Ads12815 Aspergill
12	283	87.3	64	5 AAE15373	Aae15373 Trichoder
13	283	87.3	450	4 AAB82975	Aab82975 Trichoder
14	283	87.3	451	5 AAE15371	Aae15371 Trichoder
15	283	87.3	451	8 ADS12801	Ads12801 Trichoder
16	283	87.3	451	8 ADS12804	Ads12804 Trichoder
17	277	85.5	409	7 ABO43144	Abo43144 A. thalia
18	277	85.5	409	7 ADB31925	Adb31925 Plant (A.
19	277	85.5	409	8 ADO02271	Ado02271 Thalecres
20	171.5	52.9	84	5 ABP02534	Abp02534 Human ORF
21	171.5	52.9	200	8 ADS43437	Ads43437 Bacterial
22	171.5	52.9	230	2 AAW53806	Aaw53806 Transcrip
23	171.5	52.9	230	8 ADT87049	Adt87049 Yeast Str
24	171.5	52.9	238	2 AAW53807	Aaw53807 Transcrip

RESULT 1
ADS12805
ID ADS12805 standard; protein; 349 AA.

XX ADS12805;
AC
XX
DT 16-DEC-2004 (first entry)
XX
DE Aspergillus nidulans hacA DNA binding domain.
XX
KW unfolded protein response; UPR; HAC1; PTC2;
KW unfolded protein response modulator; enzyme production; hacA; chaperone;
KW foldase; DNA binding domain.
XX

OS Emericella nidulans.

PN US2004186070-A1.

XX
PD 23-SEP-2004.

XX
PF 15-SEP-2003; 2003US-00663450.

XX
PR 24-MAR-2000; 2000US-00534692.

PR 23-MAR-2001; 2001US-00816277.

(GEMV) GENENCOR INT INC.

Penttila ME, Ward M, Wang H, Valkonen MJ, Saloheimo MLA;

WPI; 2004-707924/69.

Increasing the secretion of a heterologous protein, such as a therapeutic or an industrial enzyme, in genetically modified eukaryotic cells by inducing an elevated unfolded protein response (UPR).

Example 3; SEQ ID NO 6; 83pp; English.

The invention describes a method of increasing the secretion of a heterologous protein in a eukaryotic cell, comprising inducing an elevated unfolded protein response (UPR). Also described are: an isolated nucleic acid encoding a HAC1 protein, where the HAC1 protein induces unfolded protein response and has less than 50% similarity to yeast HAC1 protein; an isolated nucleic acid encoding a HAC1 protein, where the HAC1 protein induces unfolded protein response and where the HAC1 protein comprises a DNA binding region that has greater than 70% similarity to the DNA binding region of filamentous fungi HAC1 protein; a protein having unfolded protein response inducing activity and having greater

ALIGNMENTS

CC than 70% similarity to a fully defined amino acid sequence of 451, 349 or
CC 386, or 504 or 1232 amino acids (SEQ ID NO: 2, 4 or 16, or 12 or 14,
CC respectively) as given in the specification; a protein having an amino
CC acid sequence of SEQ ID NO: 2, 4, 8, 10, 12, 14 or 16; an isolated
CC nucleic acid encoding a PTC2 protein that modulates unfolded protein
CC response and has at least 70% similarity to a fully defined amino acid
CC sequence of 130 or 438 amino acids (SEQ ID NO: 8 or 10) as given in the
CC specification; an isolated nucleic acid encoding a IRE1 protein that
CC modulates unfolded protein response and has at least 60% similarity to a
CC fully defined amino acid sequence with SEQ ID NO: 12 or 14; a protein
CC having unfolded protein response modulating activity and having greater
CC than 70% similarity to SEQ ID NO: 8 or 10; and a cell containing a
CC heterologous nucleic acid encoding a protein having unfolded protein
CC response modulating activity and a heterologous nucleic acid encoding a
CC protein of interest to be secreted. The methods and compositions of
CC genetically manipulating cells to have an elevated unfolded protein
CC response (UPR) resulting in an increased capacity to produce secreted
CC proteins, are useful in e.g. production of therapeutic or industrial
CC enzymes. This is the amino acid sequence of *Aspergillus nidulans* HAC1
CC chaperone and foldase DNA binding domain.
XX
SQ Sequence 349 AA;

Query Match 100.0%; Score 324; DB 8; Length 349;
Best Local Similarity 100.0%; Pred. No. 9.9e-30;
Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KKPAAKRRKSGOELPVPKTNLPKRKAKTEDEKEORRIERVLRNRAAAQTSRKRKLEME 60
|||||
Db 53 KKPAAKRRKSGOELPVPKTNLPKRKAKTEDEKEORRIERVLRNRAAAQTSRKRKLEME 112

QY 61 KLES 64
|||||
Db 113 KLES 116

RESULT 2
ADSL12803
ID ADSL12803 standard; protein; 349 AA.
XX
AC ADSL12803;
XX
DT 16-DEC-2004 (first entry)
XX
DE *Aspergillus nidulans* hacA chaperone and foldase #1.
XX
KW unfolded protein response; UPR; HAC1; PTC2;
KW unfolded protein response modulator; enzyme production; hacA; chaperone;
KW foldase; enzyme.
XX
OS *Emericella nidulans*.
XX
PN US2004186070-A1.
XX
PD 23-SEP-2004.
XX
PF 15-SEP-2003; 2003US-00663450.
XX
PR 24-MAR-2000; 2000US-00534692.
PR 23-MAR-2001; 2001US-00816277.
XX
PA (GEMV) GENENCOR INT INC.
XX
PI Penttila ME, Ward M, Wang H, Valkonen MJ, Saloheimo MUA;
XX WPI; 2004-707924/69.
DR N-PSDB; ADSL12802.
DR
XX Increasing the secretion of a heterologous protein, such as a therapeutic
PT or an industrial enzyme, in genetically modified eukaryotic cells by
PT inducing an elevated unfolded protein response (UPR).
XX
PS Example 3; SEQ ID NO 4; 83pp; English.

XX The invention describes a method of increasing the secretion of a
CC heterologous protein in a eukaryotic cell, comprising inducing an
CC elevated unfolded protein response (UPR). Also described are: an isolated
CC nucleic acid encoding a HAC1 protein, where the HAC1 protein induces
CC unfolded protein response and has less than 50% similarity to yeast HAC1
CC protein; an isolated nucleic acid encoding a HAC1 protein, where the HAC1
CC protein induces unfolded protein response and where the HAC1 protein
CC comprises a DNA binding region that has greater than 70% similarity to
CC the DNA binding region of filamentous fungi HAC1 protein; a protein
CC having unfolded protein response inducing activity and having greater
CC than 70% similarity to a fully defined amino acid sequence of 451, 349 or
CC 386, or 504 or 1232 amino acids (SEQ ID NO: 2, 4 or 16, or 12 or 14,
CC respectively) as given in the specification; a protein having an amino
CC acid sequence of SEQ ID NO: 2, 4, 8, 10, 12, 14 or 16; an isolated
CC nucleic acid encoding a PTC2 protein that modulates unfolded protein
CC response and has at least 70% similarity to a fully defined amino acid
CC sequence of 130 or 438 amino acids (SEQ ID NO: 8 or 10) as given in the
CC specification; an isolated nucleic acid encoding a IRE1 protein that
CC modulates unfolded protein response and has at least 60% similarity to a
CC fully defined amino acid sequence with SEQ ID NO: 12 or 14; a protein
CC having unfolded protein response modulating activity and having greater
CC than 70% similarity to SEQ ID NO: 8 or 10; and a cell containing a
CC heterologous nucleic acid encoding a protein having unfolded protein
CC response modulating activity and a heterologous nucleic acid encoding a
CC protein of interest to be secreted. The methods and compositions of
CC genetically manipulating cells to have an elevated unfolded protein
CC response (UPR) resulting in an increased capacity to produce secreted
CC proteins, are useful in e.g. production of therapeutic or industrial
CC enzymes. This is the amino acid sequence of *Aspergillus nidulans* HAC1
CC chaperone and foldase. Note: This sequence differs from ADSL12863 given in
CC figure 8.
XX
SQ Sequence 349 AA;

Query Match 100.0%; Score 324; DB 8; Length 349;
Best Local Similarity 100.0%; Pred. No. 9.9e-30;
Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 KKPAAKRRKSGOELPVPKTNLPKRKAKTEDEKEORRIERVLRNRAAAQTSRKRKLEME 60
|||||
Db 53 KKPAAKRRKSGOELPVPKTNLPKRKAKTEDEKEORRIERVLRNRAAAQTSRKRKLEME 112

QY 61 KLES 64
|||||
Db 113 KLES 116

RESULT 3
ADSL12863
ID ADSL12863 standard; protein; 349 AA.
XX
AC ADSL12863;
XX
DT 16-DEC-2004 (first entry)
XX
DE *Aspergillus nidulans* hacA chaperone and foldase #2.
XX
KW unfolded protein response; UPR; HAC1; PTC2;
KW unfolded protein response modulator; enzyme production; HAC1; chaperone;
KW foldase; enzyme.
XX
OS *Emericella nidulans*.
XX
PN US2004186070-A1.
XX
PD 23-SEP-2004.
XX
PF 15-SEP-2003; 2003US-00663450.
XX
PR 24-MAR-2000; 2000US-00534692.
PR 23-MAR-2001; 2001US-00816277.
XX

PA (GEMV) GENENCOR INT INC.
 XX Penttila ME, Ward M, Wang H, Valkonen MJ, Saloheimo ML;A;
 XX WPI; 2004-707924/69.
 XX N-PSDB; ADS12802.
 XX Increasing the secretion of a heterologous protein, such as a therapeutic
 PT or an industrial enzyme, in genetically modified eukaryotic cells by
 PT inducing an elevated unfolded protein response (UPR).
 XX
 XX Example 3; Fig 8; 83pp; English.
 XX
 CC The invention describes a method of increasing the secretion of a
 CC heterologous protein in a eukaryotic cell, comprising inducing an
 CC elevated unfolded protein response (UPR). Also described are: an isolated
 CC nucleic acid encoding a HAC1 protein, where the HAC1 protein induces
 CC unfolded protein response and has less than 50% similarity to yeast HAC1
 CC protein; an isolated nucleic acid encoding a HAC1 protein, where the HAC1
 CC protein induces unfolded protein response and where the HAC1 protein
 CC comprises a DNA binding region that has greater than 70% similarity to
 CC the DNA binding region of filamentous fungi HAC1 protein; a protein
 CC having unfolded protein response inducing activity and having greater
 CC than 70% similarity to a fully defined amino acid sequence of 451, 349 or
 CC 386, or 504 or 1232 amino acids (SEQ ID NO: 2, 4 or 16, or 12 or 14,
 CC respectively) as given in the specification; a protein having an amino
 CC acid sequence of SEQ ID NO: 2, 4, 8, 10, 12, 14 or 16; an isolated
 CC nucleic acid encoding a PTC2 protein that modulates unfolded protein
 CC response and has at least 70% similarity to a fully defined amino acid
 CC sequence of 130 or 438 amino acids (SEQ ID NO: 8 or 10) as given in the
 CC specification; an isolated nucleic acid encoding a IRE1 protein that
 CC modulates unfolded protein response and has at least 60% similarity to a
 CC fully defined amino acid sequence with SEQ ID NO: 12 or 14; a protein
 CC having unfolded protein response modulating activity and having greater
 CC than 70% similarity to SEQ ID NO: 8 or 10; and a cell containing a
 CC heterologous nucleic acid encoding a protein having unfolded protein
 CC response modulating activity and a heterologous nucleic acid encoding a
 CC protein of interest to be secreted. The methods and compositions of
 CC genetically manipulating cells to have an elevated unfolded protein
 CC response (UPR) resulting in an increased capacity to produce secreted
 CC proteins, are useful in e.g. production of therapeutic or industrial
 CC enzymes. This is the amino acid sequence of Aspergillus nidulans HAC1
 CC chaparone and foldase. Note: This sequence differs from ADS12803 given in
 CC the sequence listing.
 XX
 SQ Sequence 349 AA;
 Query Match 100.0%; Score 324; DB 8; Length 349;
 Best Local Similarity 100.0%; Pred. No. 9.9e-30;
 Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 KKPAAKKKSGQELPVPKTNLPKRAKTEDEKEQRIERVLNRNRAAAQTSRERKLEME 60
 DB 53 KKPAAKKKSGQELPVPKTNLPKRAKTEDEKEQRIERVLNRNRAAAQTSRERKLEME 112
 QY 61 KLES 64
 DB 113 KLES 116
 RESULT 4
 ID AAB82976
 XX AAB82976 standard; protein; 350 AA.
 AC AAB82976;
 XX 11-SEP-2003 (revised)
 DT 21-DEC-2001 (first entry)
 DE Aspergillus nidulans hacA, involved in unfolded protein response.
 XX HacA; transcription factor; unfolded protein response; protein secretion.
 XX

OS Emericella nidulans.
 XX Key Location/Qualifiers
 FH Domain 53..116
 FT /label= DNA binding domain
 XX
 PN WO200172783-A2.
 XX 04-OCT-2001.
 PD XX
 XX 23-MAR-2001; 2001WO-US009401.
 PF XX
 XX 24-MAR-2000; 2000US-00534692.
 PR (GEMV) GENENCOR INT INC.
 PA Penttila ME, Ward M, Wang H, Valkonen MJ, Saloheimo ML;A;
 XX WPI; 2001-626252/72.
 DR N-PSDB; AAH26932.
 XX
 CC Increasing secretion of heterologous proteins e.g. lipase and cellulase
 PT in eukaryotic cells useful in industry to increase production and
 PT facilitate purification, by inducing an elevated unfolded protein
 PT response.
 XX
 XX Claim 54; Fig 8A-B; 89pp; English.
 XX
 CC The present sequence is that of the hacA protein of Aspergillus nidulans,
 CC as deduced from the newly isolated hacA gene (see AAH26932). HacA protein
 CC is a transcription factor involved in the unfolded protein response
 CC (UPR). The invention provides methods for increasing the secretion of a
 CC heterologous protein in a cell by inducing an elevated UPR. This can be
 CC achieved by modulating the activity of HAC1 (or hacA), PTC2 or IRE1 in
 CC the cell, e.g. by gene overexpression. The cell from which the protein is
 CC secreted can be any cell having an UPR, such as mammalian cells, insect
 CC cells, yeast and filamentous fungi. The protein of interest can be any
 CC secreted protein such as a therapeutic protein or an industrial enzyme,
 CC e.g. lipase, cellulase, endoglucanase-H, protease, carbonhydratase,
 CC reductase, oxidase, isomerase, transferase, kinase, phosphatase, alpha-
 CC amylase, glucosylase, lignocellulose hemicellulase, pectinase and
 CC ligninase (claimed). (Updated on 11-SEP-2003 to standardise OS field)
 XX
 SQ Sequence 350 AA;
 Query Match 100.0%; Score 324; DB 4; Length 350;
 Best Local Similarity 100.0%; Pred. No. 9.9e-30;
 Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 KKPAAKKKSGQELPVPKTNLPKRAKTEDEKEQRIERVLNRNRAAAQTSRERKLEME 60
 DB 54 KKPAAKKKSGQELPVPKTNLPKRAKTEDEKEQRIERVLNRNRAAAQTSRERKLEME 113
 QY 61 KLES 64
 DB 114 KLES 117
 RESULT 5
 ID AAE15372
 XX AAE15372 standard; protein; 350 AA.
 AC AAE15372;
 XX 29-AUG-2003 (revised)
 DT 07-MAR-2002 (first entry)
 DE Aspergillus nidulans hacA protein.
 XX Heterologous protein secretion; unfolded protein response; UPR; lipase;
 KW cellulase; carbohydrase; industry; purification; hacA protein.
 XX
 OS Emericella nidulans.

```

XX Key Location/Qualifiers
FH Binding-site 53..116
FT /label= DNA-binding_domain
XX
XX US2001034045-A1.
XX
XX 25-OCT-2001.
XX
XX 23-MAR-2001; 2001US-00816277.
XX
XX 24-MAR-2000; 2000US-00534692.
XX
XX (GEMV ) GENENCOR INT INC.
XX
XX Penttila ME, Ward M, Wang H, Valkonen MJ, Saloheimo MLA;
XX
XX WPI; 2002-033728/04.
XX
XX Increasing secretion of heterologous proteins e.g. lipase and cellulase
XX in eukaryotic cells useful in industry to increase production and
XX facilitate purification, by inducing an elevated unfolded protein
XX response.
XX
XX Claim 39; Fig 8; 56pp; English.
XX
XX The present invention relates to methods for increasing the secretion of
XX heterologous protein in eukaryotic cells by inducing an elevated unfolded
XX protein response (UPR). The method involves inducing the elevated UPR by
XX increasing the presence of proteins such as HAC1, HACA, PTC2 or LRE1 in
XX cells. The method and sequences are useful for increasing the secretion
XX of heterologous proteins (e.g. lipase, cellulase, carboxhydase) in
XX eukaryotic cells useful in industry to increase protein yields and to
XX facilitate purification. The present sequence is Aspergillus nidulans
XX hacA protein. (Updated on 29-AUG-2003 to standardise OS field)
XX
XX Sequence 350 AA;
XX
XX Query Match 100.0%; Score 324; DB 5; Length 350;
XX Best Local Similarity 100.0%; Pred. No. 9.9e-30;
XX Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 1 KKPAKKRKSWGQELPVPKTNLPKRKAKTEDEKEQRRIERVLRNRAAAQTSRKRKLEME 60
XX 54 KKPAKKRKSWGQELPVPKTNLPKRKAKTEDEKEQRRIERVLRNRAAAQTSRKRKLEME 113
XX
XX 61 KLES 64
XX 114 KLES 117
XX
XX RESULT 6
XX AAEE15374
XX ID AAEE15374 standard; protein; 64 AA.
XX
XX AAEE15374;
XX
XX 29-AUG-2003 (revised)
XX 07-MAR-2002 (first entry)
XX
XX Aspergillus nidulans hacA protein DNA binding domain.
XX
XX Heterologous protein secretion; unfolded protein response; UPR; lipase;
XX cellulase; carboxhydase; industry; purification; DNA binding domain;
XX hacA protein.
XX
XX Emericella nidulans.
XX
XX US2001034045-A1.
XX
XX 25-OCT-2001.
XX

```

```

PF 23-MAR-2001; 2001US-00816277.
XX
XX 24-MAR-2000; 2000US-00534692.
XX
XX (GEMV ) GENENCOR INT INC.
XX
XX Penttila ME, Ward M, Wang H, Valkonen MJ, Saloheimo MLA;
XX
XX WPI; 2002-033728/04.
XX
XX Increasing secretion of heterologous proteins e.g. lipase and cellulase
XX in eukaryotic cells useful in industry to increase production and
XX facilitate purification, by inducing an elevated unfolded protein
XX response.
XX
XX Example 3; Fig 10; 56pp; English.
XX
XX The present invention relates to methods for increasing the secretion of
XX heterologous protein in eukaryotic cells by inducing an elevated UPR by
XX protein response (UPR). The method involves inducing the elevated UPR by
XX increasing the presence of proteins such as HAC1, HACA, PTC2 or LRE1 in
XX cells. The method and sequences are useful for increasing the secretion
XX of heterologous proteins (e.g. lipase, cellulase, carboxhydase) in
XX eukaryotic cells useful in industry to increase protein yields and to
XX facilitate purification. The present sequence is Aspergillus nidulans
XX hacA protein DNA binding domain. (Updated on 29-AUG-2003 to standardise
XX OS field)
XX
XX Sequence 64 AA;
XX
XX Query Match 98.8%; Score 320; DB 5; Length 64;
XX Best Local Similarity 100.0%; Pred. No. 4.6e-30;
XX Matches 63; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 1 KKPAKKRKSWGQELPVPKTNLPKRKAKTEDEKEQRRIERVLRNRAAAQTSRKRKLEME 60
XX 2 KKPAKKRKSWGQELPVPKTNLPKRKAKTEDEKEQRRIERVLRNRAAAQTSRKRKLEME 61
XX
XX 61 KLE 63
XX 62 KLE 64
XX
XX RESULT 7
XX AAB82977
XX ID AAB82977 standard; protein; 342 AA.
XX
XX AAB82977;
XX
XX 11-SEP-2003 (revised)
XX 21-DEC-2001 (first entry)
XX
XX Aspergillus niger hacA, involved in unfolded protein response.
XX
XX HacA; transcription factor; unfolded protein response; protein secretion.
XX
XX Aspergillus awamori.
XX
XX Key Location/Qualifiers
XX Domain 45..109
XX /label= DNA binding domain
XX
XX WO200172783-A2.
XX
XX 04-OCT-2001.
XX
XX 23-MAR-2001; 2001WO-US009401.
XX
XX 24-MAR-2000; 2000US-00534692.
XX
XX (GEMV ) GENENCOR INT INC.
XX
XX Penttila ME, Ward M, Wang H, Valkonen MJ, Saloheimo MLA;
XX

```

```

XX DR WPI; 2001-626252/72.
XX DR N-PSDB; AAH26933.
XX
XX PT Increasing secretion of heterologous proteins e.g. lipase and cellulase
XX PT in eukaryotic cells useful in industry to increase production and
XX PT facilitate purification, by inducing an elevated unfolded protein
XX PT response.
XX
XX PS Claim 54; Fig 28A-C; 89pp; English.
XX
XX CC The present sequence is that of the hacA protein of Aspergillus niger
XX CC var. awamori, as deduced from hacA cDNA (see AAH26933). HacA protein is a
XX CC transcription factor involved in the unfolded protein response (UPR).
XX CC Overexpression of an inducing form of hacA enables production of higher
XX CC levels of secreted heterologous proteins in A. niger. The invention
XX CC provides methods for increasing the secretion of a heterologous protein
XX CC in a cell by inducing an elevated UPR. This can be achieved by modulating
XX CC the activity of HAC1 (or hacA), PTC2 or IRE1 in the cell, e.g. by gene
XX CC overexpression. The cell from which the protein is secreted can be any
XX CC cell having an UPR, such as mammalian cells, insect cells, yeast and
XX CC filamentous fungi. The protein of interest can be any secreted protein
XX CC such as a therapeutic protein or an industrial enzyme, e.g. lipase,
XX CC cellulase, endoglucanase-H, protease, carboxylatase, reductase, oxidase,
XX CC isomerase, transferase, kinase, phosphatase, alpha-amylase, glucoamylase,
XX CC lignocellulose hemicellulase, pectinase and ligninase (claimed). (Updated
XX CC on 11-SEP-2003 to standardise OS field)
XX
XX SQ Sequence 342 AA;
XX
XX Query Match 97.8%; Score 317; DB 4; Length 342;
XX Best Local Similarity 96.9%; Pred. No. 6.5e-29;
XX Matches 62; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
XX
OY 1 KKPAAKKRKSQGLPVPKTNLPKRKAKTEDEKEQRIERVLRNRAAAQTSRKRKLEME 60
DB 47 KKPVKKRKSQGLPVPKTNLPKRKAKTEDEKEQRIERVLRNRAAAQTSRKRKLEME 106
OY 61 KLES 64
DB 107 KLEN 110
XX
XX RESULT 8
XX ID AAEE15381 standard; protein; 342 AA.
XX AC AAEE15381;
XX DT 07-MAR-2002 (first entry)
XX DE Aspergillus niger var. awamori hacA protein #3.
XX KW Heterologous protein secretion; unfolded protein response; UPR; lipase;
XX KW cellulase; carboxylatase; industry; purification; hacA protein.
XX OS Aspergillus niger.
XX PN US2001034045-A1.
XX PD 25-OCT-2001.
XX PF 23-MAR-2001; 2001US-00816277.
XX PR 24-MAR-2000; 2000US-00534692.
XX PA (GENV) GENENCOR INT INC.
XX PI Penttila ME, Ward M, Wang H, Valkonen MJ, Saloheimo MLA;
XX DR WPI; 2002-033728/04.
XX PT Increasing secretion of heterologous proteins e.g. lipase and cellulase
XX

```

```

PT in eukaryotic cells useful in industry to increase production and
PT facilitate purification, by inducing an elevated unfolded protein
PT response.
XX
XX PS Claim 39; Fig 28; 56pp; English.
XX
XX CC The present invention relates to methods for increasing the secretion of
XX CC heterologous protein in eukaryotic cells by inducing an elevated unfolded
XX CC protein response (UPR). The method involves inducing the elevated UPR by
XX CC increasing the presence of proteins such as HAC1, HAC2, PTC2 or IRE1 in
XX CC cells. The method and sequences are useful for increasing the secretion
XX CC of heterologous proteins (e.g. lipase, cellulase, carboxylatase) in
XX CC eukaryotic cells useful in industry to increase protein yields and to
XX CC facilitate purification. The present sequence is Aspergillus niger var.
XX CC awamori hacA protein
XX
XX SQ Sequence 342 AA;
XX
XX Query Match 97.8%; Score 317; DB 5; Length 342;
XX Best Local Similarity 96.9%; Pred. No. 6.5e-29;
XX Matches 62; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
XX
OY 1 KKPAAKKRKSQGLPVPKTNLPKRKAKTEDEKEQRIERVLRNRAAAQTSRKRKLEME 60
DB 47 KKPVKKRKSQGLPVPKTNLPKRKAKTEDEKEQRIERVLRNRAAAQTSRKRKLEME 106
OY 61 KLES 64
DB 107 KLEN 110
XX
XX RESULT 9
XX ID ADS12818 standard; protein; 342 AA.
XX AC ADS12818;
XX DT 16-DEC-2004 (first entry)
XX DE Aspergillus nidulans hacA polypeptide seqid 19.
XX KW unfolded protein response; UPR; HAC1; PTC2;
XX KW unfolded protein response modulator; enzyme production; hacA; chaperone;
XX KW foldase.
XX OS Emericella nidulans.
XX PN US2004186070-A1.
XX PD 23-SEP-2004.
XX PF 15-SEP-2003; 2003US-00663450.
XX PR 24-MAR-2000; 2000US-00534692.
XX PR 23-MAR-2001; 2001US-00816277.
XX PA (GENV) GENENCOR INT INC.
XX PI Penttila ME, Ward M, Wang H, Valkonen MJ, Saloheimo MLA;
XX DR WPI; 2004-707924/69.
XX PT Increasing the secretion of a heterologous protein, such as a therapeutic
XX PT or an industrial enzyme, in genetically modified eukaryotic cells by
XX PT inducing an elevated unfolded protein response (UPR).
XX
XX PS Example 12; SEQ ID NO 19; 83pp; English.
XX
XX CC The invention describes a method of increasing the secretion of a
XX CC heterologous protein in a eukaryotic cell, comprising inducing an
XX CC elevated unfolded protein response (UPR). Also described are: an isolated
XX CC nucleic acid encoding a HAC1 protein, where the HAC1 protein induces
XX CC unfolded protein response and has less than 50% similarity to yeast HAC1
XX

```


CC protein; an isolated nucleic acid encoding a HAC1 protein, where the HAC1
 CC protein induces unfolded protein response and where the HAC1 protein to
 CC comprises a DNA binding region that has greater than 70% similarity to
 CC the DNA binding region of filamentous fungi HAC1 protein; a protein
 CC having unfolded protein response inducing activity and having greater
 CC than 70% similarity to a fully defined amino acid sequence of 451, 349 or
 CC 386, or 504 or 1232 amino acids (SEQ ID NO: 2, 4 or 16, or 12 or 14,
 CC respectively) as given in the specification; a protein having an amino
 CC acid sequence of SEQ ID NO: 2, 4, 8, 10, 12, 14 or 16; an isolated
 CC nucleic acid encoding a PTC2 protein that modulates unfolded protein
 CC response and has at least 70% similarity to a fully defined amino acid
 CC sequence of 130 or 438 amino acids (SEQ ID NO: 8 or 10) as given in the
 CC specification; an isolated nucleic acid encoding a IRE1 protein that
 CC modulates unfolded protein response and has at least 60% similarity to a
 CC fully defined amino acid sequence with SEQ ID NO: 12 or 14; a protein
 CC having unfolded protein response modulating activity and having greater
 CC than 70% similarity to SEQ ID NO: 8 or 10; and a cell containing a
 CC heterologous nucleic acid encoding a protein having unfolded protein
 CC response modulating activity and a heterologous nucleic acid encoding a
 CC protein of interest to be secreted. The methods and compositions of
 CC genetically manipulating cells to have an elevated unfolded protein
 CC response (UPR) resulting in an increased capacity to produce secreted
 CC proteins, are useful in e.g. production of therapeutic or industrial
 CC enzymes. This is the amino acid sequence of *Aspergillus nidulans* HAC1
 CC chaperone and foldase.

XX SQ Sequence 386 AA;

Query Match 97.8%; Score 317; DB 8; Length 386;
 Best Local Similarity 96.9%; Pred. No. 7.4e-29;
 Matches 62; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

OY 1 KKPAAKRSWGQELPVPKTNLPKRKAKTEDEKEORRIERVLNRAAAQTSRERKLEME 60
 DB 91 KKPVKRRKSWGQELPVPKTNLPKRKAKTEDEKEORRIERVLNRAAAQTSRERKLEME 150
 OY 61 KLES 64
 DB 151 KLEN 154

RESULT 12
 AAE15373
 ID AAE15373 standard; protein; 64 AA.

XX AC AAE15373;
 XX DT 29-AUG-2003 (revised)
 XX DT 07-MAR-2002 (first entry)
 XX DE Trichoderma reesei HAC1 protein DNA binding domain.

XX KW Heterologous protein secretion; unfolded protein response; UPR; lipase;
 KW cellulase; carboxylase; industry; purification; DNA binding domain;
 KW HAC1 protein.

XX OS Hypocrea jecorina.

XX PN US2001034045-A1.

XX PD 25-OCT-2001.

XX PF 23-MAR-2001; 2001US-00816277.

XX PR 24-MAR-2000; 2000US-00534692.

XX PA (GENV) GENENCOR INT INC.

XX PI Penttila ME, Ward M, Wang H, Valkonen MJ, Saloheimo MUA;

XX DR WPI; 2002-033728/04.

XX PT Increasing secretion of heterologous proteins e.g. lipase and cellulase

PT in eukaryotic cells useful in industry to increase production and
 PT facilitate purification, by inducing an elevated unfolded protein
 PT response.

XX Example 3; Fig 10; 56pp; English.

XX CC The present invention relates to methods for increasing the secretion of
 CC heterologous protein in eukaryotic cells by inducing an elevated unfolded
 CC protein response (UPR). The method involves inducing the elevated UPR by
 CC increasing the presence of proteins such as HAC1, HACA, PTC2 or IRE1 in
 CC cells. The method and sequences are useful for increasing the secretion
 CC of heterologous proteins (e.g. lipase, cellulase, carboxylase) in
 CC eukaryotic cells useful in industry to increase protein yields and to
 CC facilitate purification. The present sequence is *Trichoderma reesei* HAC1
 CC protein DNA binding domain. (Updated on 29-AUG-2003 to standardise OS
 CC field)

XX SQ Sequence 64 AA;

Query Match 87.3%; Score 283; DB 5; Length 64;
 Best Local Similarity 87.3%; Pred. No. 1.1e-25;
 Matches 55; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

OY 1 KKPAAKRSWGQELPVPKTNLPKRKAKTEDEKEORRIERVLNRAAAQTSRERKLEME 60
 DB 2 KKPVKRRKSWGQELPVPKTNLPKRKAKTEDEKEORRIERVLNRAAAQTSRERKLEME 61
 OY 61 KLE 63
 DB 62 ALE 64

RESULT 13
 AAB82975
 ID AAB82975 standard; protein; 450 AA.

XX AC AAB82975;

XX DT 11-SEP-2003 (revised)

XX DT 21-DEC-2001 (first entry)

XX DE Trichoderma reesei HAC1, involved in unfolded protein response.

XX KW HAC1; transcription factor; unfolded protein response; protein secretion.

XX OS Hypocrea jecorina.

XX FH Key Location/Qualifiers

XX FT Domain 84..147

XX FT /label= DNA binding domain

XX PN WO200172783-A2.

XX PD 04-OCT-2001.

XX PF 23-MAR-2001; 2001WO-US009401.

XX PR 24-MAR-2000; 2000US-00534692.

XX PA (GENV) GENENCOR INT INC.

XX PI Penttila ME, Ward M, Wang H, Valkonen MJ, Saloheimo MUA;

XX DR WPI; 2001-626252/72.

XX DR N-PSDB; AAB26931.

XX PT Increasing secretion of heterologous proteins e.g. lipase and cellulase
 PT in eukaryotic cells useful in industry to increase production and
 PT facilitate purification, by inducing an elevated unfolded protein
 PT response.

XX PS Claim 54; Fig 7A-B; 89pp; English.

CC 386, or 504 or 1232 amino acids (SEQ ID NO: 2, 4 or 16, or 12 or 14,
CC respectively) as given in the specification; a protein having an amino
CC acid sequence of SEQ ID NO: 2, 4, 8, 10, 12, 14 or 16; an isolated
CC nucleic acid encoding a PTC2 protein that modulates unfolded protein
CC response and has at least 70% similarity to a fully defined amino acid
CC sequence of 130 or 438 amino acids (SEQ ID NO: 8 or 10) as given in the
CC specification; an isolated nucleic acid encoding a IRE1 protein that
CC modulates unfolded protein response and has at least 60% similarity to a
CC fully defined amino acid sequence with SEQ ID NO: 12 or 14; a protein
CC having unfolded protein response modulating activity and having greater
CC than 70% similarity to SEQ ID NO: 8 or 10; and a cell containing a
CC heterologous nucleic acid encoding a protein having unfolded protein
CC response modulating activity and a heterologous nucleic acid encoding a
CC protein of interest to be secreted. The methods and compositions of
CC genetically manipulating cells to have an elevated unfolded protein
CC response (UPR) resulting in an increased capacity to produce secreted
CC proteins, are useful in e.g. production of therapeutic or industrial
CC enzymes. This is the amino acid sequence of Trichoderma reesei HAC1
CC chaperone and foldase.

XX

SQ Sequence 451 AA;

Query Match 87.3%; Score 283; DB 8; Length 451;
Best Local Similarity 87.3%; Pred. No. 9.1e-25;
Matches 55; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

QY 1 KKPAAKRRKSWGQELVPVPTNLPVRKAETDEKEQRRIRVLRNRAAAQTSSRRKRLEME 60
DB 85 KKPVKRRKSWGQVLPEPTNLPVRKAETDEKEQRRIRVLRNRAAQSSRRKRLEVE 144
QY 61 KLE 63
DB 145 ALR 147

Search completed: November 22, 2005, 15:43:40
Job time : 169 secs

THIS PAGE BLANK (USPTO)

Result No.	Score	Query Match	Length	DB	ID	Description
1	307	95.0	409	2	US-09-533-029-104	Sequence 104, App
2	174	53.9	248	2	US-09-248-796A-18840	Sequence 18840, A
3	99	30.7	143	2	US-09-640-211A-1930	Sequence 1930, Ap
4	89	27.6	23	2	US-09-831-642-72	Sequence 72, Appl
5	89	27.6	371	2	US-09-148-545-259	Sequence 259, App
6	89	27.6	371	2	US-09-621-011-259	Sequence 259, App
7	86	26.6	102	2	US-09-640-211A-1046	Sequence 1046, Ap
8	82.5	25.5	611	2	US-09-216-393B-81	Sequence 81, Appl
9	81	25.1	212	2	US-09-489-039A-13074	Sequence 13074, A
10	80.5	24.9	385	2	US-09-247-155-113	Sequence 113, App
11	80.5	24.9	385	2	US-09-513-998C-14	Sequence 14, Appl
12	80.5	24.9	395	2	US-09-471-276-14	Sequence 14, Appl
13	80.5	24.9	395	2	US-09-903-190-113	Sequence 113, App
14	80	24.8	63	2	US-09-107-433-3674	Sequence 3674, Ap
15	78	24.1	551	2	US-08-796-899-29	Sequence 29, Appl
16	77.5	24.0	501	2	US-09-949-016-11281	Sequence 11281, A
17	77	23.8	284	2	US-09-949-016-10789	Sequence 10789, A
18	76.5	23.7	351	2	US-09-870-089B-2	Sequence 2, Appli
19	76.5	23.7	362	2	US-09-949-016-10923	Sequence 10923, A
20	75.5	23.4	463	2	US-09-345-236B-2	Sequence 2, Appli
21	75	23.2	170	2	US-09-248-796A-27838	Sequence 27838, A
22	73.5	22.8	260	2	US-09-538-092-950	Sequence 950, App
23	73.5	22.8	338	1	US-08-218-686-2	Sequence 2, Appli
24	73.5	22.8	338	2	US-08-460-242-2	Sequence 2, Appli
25	73.5	22.8	2058	2	US-09-949-016-6835	Sequence 6835, Ap
26	73.5	22.8	2111	2	US-09-949-016-10199	Sequence 10199, A
27	73	22.6	325	2	US-09-267-031-14	Sequence 14, Appl

```

; Patent No. 6747137
; GENERAL INFORMATION:
; APPLICANT: Keith Weinstein et al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA ALBICAN
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.132
; CURRENT APPLICATION NUMBER: US/09/248,796A
; PRIOR FILING DATE: 1999-02-12
; PRIOR FILING DATE: 1998-02-13
; PRIOR FILING DATE: 1998-02-13
; PRIOR FILING DATE: 1998-02-13
; NUMBER OF SEQ ID NOS: 28208
; SEQ ID NO 18840
; LENGTH: 248
; TYPE: PRT
; ORGANISM: Candida albicans
; FEATURE:
; NAME/KEY: UNSURE
; LOCATION: (221)
; OTHER INFORMATION: Identity of amino acid sequences at the above locations are unknown
US-09-248-796A-18840

Query Match      53.9%; Score 174; DB 2; Length 248;
Best Local Similarity 71.7%; Pred. No. 2.6e-12;
Matches 33; Conservative 8; Mismatches 5; Indels 0; Gaps 0;

QY 19 KTNLPKRAKTEDEKEQRRVERVLNRRRAAQSRRERKEVEALE 64
Db 72 KSTLPKRAKTEDEKEQRRVERVLNRRRAAQSRRERKEVEALE 117

RESULT 3
US-09-640-211A-1930
; Sequence 1930, Application US/09640211A
; Patent No. 6833446
; GENERAL INFORMATION:
; APPLICANT: Wood, Marion
; APPLICANT: Shenk, Michael A.
; APPLICANT: McGrath, Annette
; APPLICANT: Glenn, Matthew
; TITLE OF INVENTION: Compositions and Methods for the
; TITLE OF INVENTION: Modification of Gene Transcription
; FILE REFERENCE: 11000.1021C1U
; CURRENT APPLICATION NUMBER: US/09/640,211A
; CURRENT FILING DATE: 2000-08-16
; NUMBER OF SEQ ID NOS: 2368
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1930
; LENGTH: 143
; TYPE: PRT
; ORGANISM: Pinus radiata
US-09-640-211A-1930

Query Match      30.7%; Score 99; DB 2; Length 143;
Best Local Similarity 47.6%; Pred. No. 0.00056;
Matches 20; Conservative 9; Mismatches 13; Indels 0; Gaps 0;

QY 23 PPKRAKTEDEKEQRRVERVLNRRRAAQSRRERKEVEALE 64
Db 35 PPKRAADLNAEORREARAHNRNIAAQNRSRDKRKAQTYME 76

RESULT 4
US-09-831-642-72
; Sequence 72, Application US/09831642
; Patent No. 6635751
; GENERAL INFORMATION:
; APPLICANT: HAZE, Kyosuke et al.
; TITLE OF INVENTION: ENDOPLASMIC RETICULUM STRESS TRANSCRIPTION FACTORS ATF6 AND CREB-
; FILE REFERENCE: 1422-0474P
; CURRENT APPLICATION NUMBER: US/09/831,642
; CURRENT FILING DATE: 2001-05-11
```

```

; NUMBER OF SEQ ID NOS: 83
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 72
; LENGTH: 23
; TYPE: PRT
; ORGANISM: Yeast HacIp
US-09-831-642-72

Query Match      27.6%; Score 89; DB 2; Length 23;
Best Local Similarity 77.3%; Pred. No. 0.001;
Matches 17; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 37 RRVRLNRRRAAQSRRERKRL 58
Db 1 RRIERILNRRRAAQSRRERKRL 22

RESULT 5
US-09-148-545-259
; Sequence 259, Application US/09148545
; Patent No. 6590075
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: 70 Human Secreted Proteins
; FILE REFERENCE: PZ001P1
; CURRENT APPLICATION NUMBER: US/09/148,545
; CURRENT FILING DATE: 1998-09-04
; EARLIER APPLICATION NUMBER: PCT/US98/04482
; EARLIER FILING DATE: 1998-03-06
; EARLIER APPLICATION NUMBER: 60/040,162
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,333
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/038,621
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,161
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,626
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,334
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,336
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/040,163
; EARLIER FILING DATE: 1997-03-07
; EARLIER APPLICATION NUMBER: 60/047,615
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,600
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,597
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,502
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,633
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,583
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,617
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,618
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,503
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,592
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,581
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,584
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,500
; EARLIER FILING DATE: 1997-05-23
; EARLIER APPLICATION NUMBER: 60/047,587
; EARLIER FILING DATE: 1997-05-23
```

1	EARLIER APPLICATION NUMBER: 60/047,492
2	EARLIER FILING DATE: 1997-05-23
3	EARLIER APPLICATION NUMBER: 60/047,598
4	EARLIER FILING DATE: 1997-05-23
5	EARLIER APPLICATION NUMBER: 60/047,613
6	EARLIER FILING DATE: 1997-05-23
7	EARLIER APPLICATION NUMBER: 60/047,582
8	EARLIER FILING DATE: 1997-05-23
9	EARLIER APPLICATION NUMBER: 60/047,596
10	EARLIER FILING DATE: 1997-05-23
11	EARLIER APPLICATION NUMBER: 60/047,612
12	EARLIER FILING DATE: 1997-05-23
13	EARLIER APPLICATION NUMBER: 60/047,632
14	EARLIER FILING DATE: 1997-05-23
15	EARLIER APPLICATION NUMBER: 60/047,601
16	EARLIER FILING DATE: 1997-05-23
17	EARLIER APPLICATION NUMBER: 60/043,580
18	EARLIER FILING DATE: 1997-04-11
19	EARLIER APPLICATION NUMBER: 60/043,568
20	EARLIER FILING DATE: 1997-04-11
21	EARLIER APPLICATION NUMBER: 60/043,314
22	EARLIER FILING DATE: 1997-04-11
23	EARLIER APPLICATION NUMBER: 60/043,569
24	EARLIER FILING DATE: 1997-04-11
25	EARLIER APPLICATION NUMBER: 60/043,311
26	EARLIER FILING DATE: 1997-04-11
27	EARLIER APPLICATION NUMBER: 60/043,671
28	EARLIER FILING DATE: 1997-04-11
29	EARLIER APPLICATION NUMBER: 60/043,674
30	EARLIER FILING DATE: 1997-04-11
31	EARLIER APPLICATION NUMBER: 60/043,669
32	EARLIER FILING DATE: 1997-04-11
33	EARLIER APPLICATION NUMBER: 60/043,312
34	EARLIER FILING DATE: 1997-04-11
35	EARLIER APPLICATION NUMBER: 60/043,315
36	EARLIER FILING DATE: 1997-04-11
37	EARLIER APPLICATION NUMBER: 60/048,974
38	EARLIER FILING DATE: 1997-06-06
39	EARLIER APPLICATION NUMBER: 60/056,886
40	EARLIER FILING DATE: 1997-08-22
41	EARLIER APPLICATION NUMBER: 60/056,877
42	EARLIER FILING DATE: 1997-08-22
43	EARLIER APPLICATION NUMBER: 60/056,889
44	EARLIER FILING DATE: 1997-08-22
45	EARLIER APPLICATION NUMBER: 60/056,893
46	EARLIER FILING DATE: 1997-08-22
47	EARLIER APPLICATION NUMBER: 60/056,662
48	EARLIER FILING DATE: 1997-08-22
49	EARLIER APPLICATION NUMBER: 60/056,630
50	EARLIER FILING DATE: 1997-08-22
51	EARLIER APPLICATION NUMBER: 60/056,878
52	EARLIER FILING DATE: 1997-08-22
53	EARLIER APPLICATION NUMBER: 60/056,662
54	EARLIER FILING DATE: 1997-08-22
55	EARLIER APPLICATION NUMBER: 60/056,872
56	EARLIER FILING DATE: 1997-08-22
57	EARLIER APPLICATION NUMBER: 60/056,882
58	EARLIER FILING DATE: 1997-08-22
59	EARLIER APPLICATION NUMBER: 60/056,637
60	EARLIER FILING DATE: 1997-08-22
61	EARLIER APPLICATION NUMBER: 60/056,903
62	EARLIER FILING DATE: 1997-08-22
63	EARLIER APPLICATION NUMBER: 60/056,888
64	EARLIER FILING DATE: 1997-08-22
65	EARLIER APPLICATION NUMBER: 60/056,879
66	EARLIER FILING DATE: 1997-08-22
67	EARLIER APPLICATION NUMBER: 60/056,880
68	EARLIER FILING DATE: 1997-08-22
69	EARLIER APPLICATION NUMBER: 60/056,894
70	EARLIER FILING DATE: 1997-08-22
71	EARLIER APPLICATION NUMBER: 60/056,911

```

, EARLIER FILING DATE: 1997-08-22
, EARLIER APPLICATION NUMBER: 60/056,636
, EARLIER FILING DATE: 1997-08-22
, EARLIER APPLICATION NUMBER: 60/056,874
, EARLIER FILING DATE: 1997-08-22
, EARLIER APPLICATION NUMBER: 60/056,910
, EARLIER FILING DATE: 1997-08-22
, EARLIER APPLICATION NUMBER: 60/056,864
, EARLIER FILING DATE: 1997-08-22
, EARLIER APPLICATION NUMBER: 60/056,631
, EARLIER FILING DATE: 1997-08-22
, EARLIER APPLICATION NUMBER: 60/056,845
, EARLIER FILING DATE: 1997-08-22
, EARLIER APPLICATION NUMBER: 60/056,892
, EARLIER FILING DATE: 1997-08-22
, EARLIER APPLICATION NUMBER: 60/047,595
, EARLIER FILING DATE: 1997-05-23
, EARLIER APPLICATION NUMBER: 60/057,761
, EARLIER FILING DATE: 05-Sep-1997
, EARLIER APPLICATION NUMBER: 60/047,599
, EARLIER FILING DATE: 1997-05-23
, EARLIER APPLICATION NUMBER: 60/047,588
, EARLIER FILING DATE: 1997-05-23
, EARLIER APPLICATION NUMBER: 60/047,585
, EARLIER FILING DATE: 1997-05-23
, EARLIER APPLICATION NUMBER: 60/047,586
, EARLIER FILING DATE: 1997-05-23
, EARLIER APPLICATION NUMBER: 60/047,590
, EARLIER FILING DATE: 1997-05-23
, EARLIER APPLICATION NUMBER: 60/047,594
, EARLIER FILING DATE: 1997-05-23
, EARLIER APPLICATION NUMBER: 60/047,589
, EARLIER FILING DATE: 1997-05-23
, EARLIER APPLICATION NUMBER: 60/047,593
, EARLIER FILING DATE: 1997-05-23
, EARLIER APPLICATION NUMBER: 60/047,614
, EARLIER FILING DATE: 1997-05-23
, EARLIER APPLICATION NUMBER: 60/043,578
, EARLIER FILING DATE: 1997-04-11
, EARLIER APPLICATION NUMBER: 60/043,576
, EARLIER FILING DATE: 1997-04-11
, EARLIER APPLICATION NUMBER: 60/047,501
, EARLIER FILING DATE: 1997-05-23
, EARLIER APPLICATION NUMBER: 60/043,670
, EARLIER FILING DATE: 1997-04-11
, EARLIER APPLICATION NUMBER: 60/056,632
, EARLIER FILING DATE: 1997-08-22
, EARLIER APPLICATION NUMBER: 60/056,664
, EARLIER FILING DATE: 1997-08-22
, EARLIER APPLICATION NUMBER: 60/056,876
, EARLIER FILING DATE: 1997-08-22
, EARLIER APPLICATION NUMBER: 60/056,881
, EARLIER FILING DATE: 1997-08-22
, EARLIER APPLICATION NUMBER: 60/056,909
, EARLIER FILING DATE: 1997-08-22
, EARLIER APPLICATION NUMBER: 60/056,875
, EARLIER FILING DATE: 1997-08-22
, EARLIER APPLICATION NUMBER: 60/056,862
, EARLIER FILING DATE: 1997-08-22
, EARLIER APPLICATION NUMBER: 60/056,887
, EARLIER FILING DATE: 1997-08-22
, EARLIER APPLICATION NUMBER: 60/056,908
, EARLIER FILING DATE: 1997-08-22
, EARLIER APPLICATION NUMBER: 60/048,964
, EARLIER FILING DATE: 1997-06-06
, EARLIER APPLICATION NUMBER: 60/057,650
, EARLIER FILING DATE: 1997-09-05
, EARLIER APPLICATION NUMBER: 60/056,884
, EARLIER FILING DATE: 1997-08-22
, NUMBER OF SEQ ID NOS: 280
, SOFTWARE: PatentIn ver. 2.0
, SEQ ID NO 259
, LENGTH: 371

```


APPLICANT: Bougueleret, Lydie
; TITLE OF INVENTION: Complementary DNAs
; FILE REFERENCE: GENSET.021A
; CURRENT APPLICATION NUMBER: US/09/247,155A
; CURRENT FILING DATE: 1999-02-09
; EARLIER APPLICATION NUMBER: 60/074,121
; EARLIER FILING DATE: 1998-02-09
; EARLIER APPLICATION NUMBER: 60/081,563
; EARLIER FILING DATE: 1998-04-13
; EARLIER APPLICATION NUMBER: 60/096,116
; EARLIER FILING DATE: 1998-08-10
; EARLIER APPLICATION NUMBER: 60/099,273
; EARLIER FILING DATE: 1998-10-04
; NUMBER OF SEQ ID NOS: 182
; SOFTWARE: Patent.pm
; SEQ ID NO 113
; LENGTH: 395
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SIGNAL
; LOCATION: -310...-1
US-09-247-155-113

Query Match 24.9%; Score 80.5; DB 2; Length 395;
Best Local Similarity 27.6%; Pred. No. 0.23; Mismatches 13; Indels 23; Gaps 1;
Matches 21; Conservative 13

QY 12 GOVLPEPTNLPPRKRAKTEDEKEQ-----RRVERVLNRRA 48
DB 171 GTVAPVCTLLPCQTLFTDEKRLGQGVSLPSHPLTKAERVLKVKRRKIRKQS 230
QY 49 AQSSRRKRLEVEALE 64
DB 231 AQSSRRRKKEYIDGLE 246

RESULT 11
US-09-513-999C-14
; Sequence 14, Application US/09513999C
; Patent No. 6783961
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Duclert, A.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
; FILE REFERENCE: 59 US2 REG
; CURRENT APPLICATION NUMBER: US/09/513,999C
; CURRENT FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/122,487
; PRIOR FILING DATE: 1999-02-26
; NUMBER OF SEQ ID NOS: 36681
; SOFTWARE: Patent.pm
; SEQ ID NO 14
; LENGTH: 395
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SIGNAL
; LOCATION: -310...-1
; OTHER INFORMATION: score 4.4
; OTHER INFORMATION: seq VLLIFSLALIL/PS
US-09-513-999C-14

Query Match 24.9%; Score 80.5; DB 2; Length 395;
Best Local Similarity 27.6%; Pred. No. 0.23; Mismatches 13; Indels 23; Gaps 1;
Matches 21; Conservative 13

QY 12 GOVLPEPTNLPPRKRAKTEDEKEQ-----RRVERVLNRRA 48
DB 171 GTVAPVCTLLPCQTLFTDEKRLGQGVSLPSHPLTKAERVLKVKRRKIRKQS 230

QY 49 AQSSRRKRLEVEALE 64
DB 231 AQSSRRRKKEYIDGLE 246
RESULT 12
US-09-471-276-14
; Sequence 14, Application US/09471276
; Patent No. 6822072
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, J.B.
; APPLICANT: Duclert A.
; APPLICANT: Giordano, J.Y.
; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.
; Patent No. 6822072
; FILE REFERENCE: GENSET.025CPI
; CURRENT APPLICATION NUMBER: US/09/471,276
; CURRENT FILING DATE: 1999-12-21
; EARLIER APPLICATION NUMBER: 09/057,719
; EARLIER FILING DATE: 1998-04-09
; EARLIER APPLICATION NUMBER: 09/069,047
; EARLIER FILING DATE: 1998-04-28
; EARLIER APPLICATION NUMBER: PCT/IB99/00712
; EARLIER FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 1622
; SOFTWARE: Patent.pm
; SEQ ID NO 14
; LENGTH: 395
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SIGNAL
; LOCATION: -310...-1
; OTHER INFORMATION: score 4.4
; OTHER INFORMATION: seq VLLIFSLALIL/PS
US-09-471-276-14

Query Match 24.9%; Score 80.5; DB 2; Length 395;
Best Local Similarity 27.6%; Pred. No. 0.23; Mismatches 21; Conservative 13; Indels 23; Gaps 1;
Matches 21

QY 12 GOVLPEPTNLPPRKRAKTEDEKEQ-----RRVERVLNRRA 48
DB 171 GTVAPVCTLLPCQTLFTDEKRLGQGVSLPSHPLTKAERVLKVKRRKIRKQS 230
QY 49 AQSSRRKRLEVEALE 64
DB 231 AQSSRRRKKEYIDGLE 246

RESULT 13
US-09-903-190-113
; Sequence 113, Application US/09903190
; Patent No. 6936592
; GENERAL INFORMATION:
; APPLICANT: Dumas Milne Edwards, Jean-Baptiste
; APPLICANT: Duclert, Aymeric
; APPLICANT: Bougueleret, Lydie
; TITLE OF INVENTION: Complementary DNAs
; FILE REFERENCE: GENSET.021A
; CURRENT APPLICATION NUMBER: US/09/903,190
; CURRENT FILING DATE: 2001-07-11
; PRIOR APPLICATION NUMBER: US/09/247,155A
; PRIOR FILING DATE: 1999-02-09
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/074,121
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-02-09
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/081,563
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-04-13
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/096,116
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-08-10
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 60/099,273
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-10-04
; NUMBER OF SEQ ID NOS: 182

```

; SOFTWARE: Patent.pm
; SEQ ID NO 113
; LENGTH: 395
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: SIGNAL
; LOCATION: -310...-1
US-09-903-190-113

Query Match      24.9%; Score 80.5; DB 2; Length 395;
Best Local Similarity 27.6%; Pred. No. 0.23;
Matches 21; Conservative 13; Mismatches 19; Indels 23; Gaps 1;

QY 12 GQVLPPEKTNLPPRKAKTEDEKEQ-----RRVERVLNRRRA 48
DB 171 GTVAPVPCVCTLLPCQTLFLTDEKRLIGQGVSLPSHLPLTKAEERVLKKVREKIRNKQS 230
QY 49 AQSSRRERKRLEVEALE 64
DB 231 AQDSRRRRKKKEYIDGLE 246

RESULT 14
US-09-107-433-3674
; Sequence 3674, Application US/09107433
; Patent No. 6800744
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID
; SEQUENCES RELATING TO STREPTOCOCCUS PNEUMONIAE
; THERAPEUTICS
; NUMBER OF SEQUENCES: 5206
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Walcham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: <Unknown>
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: <Unknown>
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,433
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/ 085131
; FILING DATE: May 12, 1998
; APPLICATION NUMBER: 60/051553
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-011
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 3674:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 63 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Streptococcus pneumoniae
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (B) LOCATION 1...63
; SEQUENCE DESCRIPTION: SEQ ID NO: 3674:

```

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 22, 2005, 15:34:11 ; Search time 24 Seconds
(without alignments)
256.578 Million cell updates/sec

Title: US-10-663-450-5_COPY_84_147

Perfect score: 323

Sequence: 1 EKFPVKRRKSGQVLPEPKT.....NRRAAQSSRRKRLEVEALE 64

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: PIR1:*
- 2: PIR2:*
- 3: PIR3:*
- 4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	175	54.2	238	2 S78571	transcription fact
2	106	32.8	168	2 T05922	bZIP protein HY5
3	91	28.2	322	2 T08592	TCAGC-motif-bindin
4	91	28.2	326	2 T08591	TCAGC-motif bindin
5	90.5	28.0	688	2 T32750	hypothetical prote
6	88	27.2	322	2 T12093	TCAGC-motif bindin
7	87.5	27.1	176	2 B90087	hypothetical prote
8	87	26.9	600	2 T00759	probable G-box bin
9	86	26.6	360	2 T03373	hypothetical prote
10	81	25.1	445	2 T50372	probable zutotin [i
11	80.5	24.9	486	2 JC4028	activating transcr
12	79.5	24.6	2052	2 T18519	myosin X - bovine
13	79	24.5	381	2 S26812	transcription fact
14	79	24.5	672	2 T21469	hypothetical prote
15	78.5	24.3	232	2 S42392	G-box-binding prot
16	78.5	24.3	242	2 S05453	transcription fact
17	78	24.1	424	2 T10985	regulator protein
18	77.5	24.0	246	2 T12585	Dc3 promoter-bindi
19	77.5	24.0	483	2 S12741	transcription fact
20	77.5	24.0	505	1 S05380	transcription fact
21	77	23.8	1089	2 T36653	protein kinase, tr
22	76.5	23.7	144	2 T14796	hypothetical prote
23	76.5	23.7	313	2 A34785	DNA-binding protei
24	76.5	23.7	351	2 A45377	transcription fact
25	76.5	23.7	358	2 C42026	cyclic AMP respons
26	76.5	23.7	389	1 A39429	CAMP response elem
27	76.5	23.7	448	2 A42026	CAMP response elem
28	76.5	23.7	456	2 B42026	cyclic AMP respons
29	76	23.5	502	2 T20130	hypothetical prote

30	76	23.5	521	2 S06219	colicin E1 - Shige
31	76	23.5	938	2 T20125	hypothetical prote
32	75.5	23.4	315	2 S20883	G-box-binding fact
33	75.5	23.4	315	2 G85433	G-box-binding fact
34	75.5	23.4	605	2 T02350	hypothetical prote
35	75	23.2	267	2 JC4857	hepatocarcinogenes
36	75	23.2	267	2 S51307	G-box binding fact
37	75	23.2	433	2 JC1230	DNA-binding protei
38	75	23.2	433	2 JC1230	hypothetical prote
39	75	23.2	761	2 T00940	myosin X - mouse
40	74.5	23.1	313	2 S66312	G-box binding fact
41	74.5	23.1	349	2 A41349	histone-specific t
42	74.5	23.1	349	2 S77570	transcription fact
43	74.5	23.1	452	2 H96710	hypothetical prote
44	74	22.9	331	2 T26807	hypothetical prote
45	74	22.9	333	2 T26808	hypothetical prote

ALIGNMENTS

RESULT 1

S78571 transcription factor HAC1 - yeast (Saccharomyces cerevisiae)

N;Alternate names: protein YFL031w

C;Species: Saccharomyces cerevisiae

C;Date: 13-Sep-1998 #sequence revision 13-Sep-1998 #text_change 09-Jul-2004

C;Accession: S78571; S56223; S53578

R;Murakami, Y.

submitted to the Protein Sequence Database, January 1998

A;Reference number: S78570

A;Accession: S78571

A;Molecule type: DNA

A;Residues: 1-238 <MUR>

A;Cross-references: UNIPROT:P41546; UNIPARC:UPI0000157F61; EMBL:D50617; MIPS:YFL031w

A;Note: this is a revision to the sequence from reference S56186

R;Murakami, Y.; Naitou, M.; Hagiwara, H.; Shibata, T.; Ozawa, M.; Sasanuma, S.I.; Sasanuma, S.I.

submitted to the EMBL Data Library, May 1995

A;Description: Analysis of the nucleotide sequence of chromosome VI from Saccharomyces cerevisiae

A;Reference number: S56186

A;Accession: S56223

A;Molecule type: DNA

A;Residues: 1-191, RWQTHSATI, <MUR>

A;Cross-references: UNIPARC:UPI0000179B60; EMBL:D50617; NID:9836685; PID:9836723; MIPS:YFL031w

A;Note: this sequence has been revised in reference S78570

A;Note: this was believed to be the complete sequence of YFL031w

R;Nojima, H.; Leem, S.H.; Araki, H.; Sakai, A.; Nakashima, N.; Kanaoka, Y.; Ono, Y.

Nucleic Acids Res. 22, 5279-5288, 1994

A;Title: Hac1: a novel yeast bZIP protein binding to the CRE motif is a multicopy suppressor of the bZIP protein

A;Reference number: S53578; MUID:95116316; PMID:7816617

A;Accession: S53578

A;Molecule type: DNA

A;Residues: 1-142, 'RLCRPRVCRFRVGRDFMGAECLRRKMYQSRRLPYTI', 183-220, 'AVITWTRKIQ' <NOJ>

A;Cross-references: UNIPARC:UPI000012C042; GB:D26506; NID:9633122; PID:BAA05513.1; PID:9633122

C;Genetics: A;Gene: SGD:HAC1; IRE2; ERN4

A;Cross-references: SGD:S0001863; MIPS:YFL031w

A;Map position: 6L

A;Introns: 221/1

C;Keywords: DNA binding; nucleus; transcription factor

Query Match 54.2%; Score 175; DB 2; Length 238;

Best Local Similarity 71.7%; Pred. No. 6.5e-09;

Matches 33; Conservative 8; Mismatches 5; Indels 0; Gaps 0;

QY 19 KTNLPKPKAKTEDEKEQRRVERVLNRRRAAQSRRKRLEVEALE 64

Db 23 KSTLPKPKAKTKKEQRRVERVLNRRRAAQSRRKRLEVEALE 68

RESULT 2

T50922

bZIP protein HY5 [imported] - Arabidopsis thaliana

C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 21-Jul-2000 #sequence_revision 21-Jul-2000 #text_change 09-Jul-2004
C;Accession: T50922
R;Oyama, T.; Shimura, Y.; Okada, K.
submitted to the EMBL Data Library, July 1997
A;Description: The Arabidopsis HY5 gene encodes a bZIP protein that regulates stimulus-induced transcriptional activation of the HY5 promoter

A;Reference number: Z25271
A;Accession: T50922
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-168 <OVA>
A;Cross-references: UNIPROT:Q24646; UNIPARC:UPI000012CFCA; EMBL:AB005295; PIDN:BAA21116
A;Experimental source: Landsberg erecta

Query Match 32.0%; Score 106; DB 2; Length 168;
Best Local Similarity 43.4%; Pred.No. 0.0078;
Matches 23; Conservative 10; Mismatches 20; Indels 0; Gaps 0;

QY 12 GQVLPKTNLPKPKAKTEDEKEQRVERVLNRRAAQSSRRKRLEVALE 64
Db 65 GOERTATGVSQQRKGRTPAEKENRLKLLRNVRSAQAQARRKKAYLSELE 117

RESULT 3
T08592
TGACG-motif-binding protein STF2 - soybean
C;Species: Glycine max (soybean)
C;Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 31-Dec-2004
C;Accession: T08592
R;Cheong, Y.H.; Yoo, C.M.; Park, J.M.; Ryu, G.R.; Goekjian, V.H.; Nagao, R.T.; Key, J.L.
submitted to the EMBL Data Library, September 1995
A;Description: STF1 is a novel TGACG-binding factor with a zinc-finger motif and a bZIP
A;Reference number: Z16445
A;Accession: T08592
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-322 <CHE>
A;Cross-references: UNIPROT:Q39896; UNIPARC:UPI00000A61C9; EMBL:L28004; NID:g986966; PIDN:U000000000
A;Experimental source: strain Williams; hypocotyl
C;Genetics:
A;Gene: STF2
C;Superfamily: TGACG-motif-binding transcription factor

Query Match 28.2%; Score 91; DB 2; Length 322;
Best Local Similarity 45.0%; Pred.No. 0.31;
Matches 18; Conservative 12; Mismatches 10; Indels 0; Gaps 0;

QY 25 RKRAKTEDEKEQRVERVLNRRAAQSSRRKRLEVALE 64
Db 234 KKGRSPADKESKRLKLLRNVRSAQAQARRKKAYLIDLE 273

RESULT 4
T08591
TGACG-motif binding protein STF1 - soybean
C;Species: Glycine max (soybean)
C;Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 31-Dec-2004
C;Accession: T08591
R;Cheong, Y.H.; Yoo, C.M.; Park, J.M.; Ryu, G.R.; Goekjian, V.H.; Nagao, R.T.; Key, J.L.
submitted to the EMBL Data Library, September 1995
A;Description: STF1 is a novel TGACG-binding factor with a zinc-finger motif and a bZIP
A;Reference number: Z16445
A;Accession: T08591
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-326 <CHE>
A;Cross-references: UNIPROT:Q39895; UNIPARC:UPI00000A6233; EMBL:L28003; NID:g2934883; PIDN:U000000000
A;Experimental source: strain Williams; hypocotyl
C;Superfamily: TGACG-motif-binding transcription factor

Query Match 28.2%; Score 91; DB 2; Length 326;
Best Local Similarity 45.0%; Pred.No. 0.32;
Matches 18; Conservative 12; Mismatches 10; Indels 0; Gaps 0;

Nature 410, 1091-1096, 2001
A:Title: The highly reduced genome of an enslaved algal nucleus.

A:Reference number: A99082; MUID:11323671; PMID:11323671

A:Accession: B90087

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-176 <DOU>

A:Cross-references: UNIPROT:Q98RX1; UNIPARC:UPI000009BF82; GB:AF165818; NID:gl3794454; F

C:Genetics:

A:Gene: orf176

A:Map position: 1

A:Genome: nucleomorph

C:Keywords: nucleomorph

Query Match 27.1%; Score 87.5; DB 2; Length 176;
Best Local Similarity 45.0%; Pred. No. 0.38;

Matches 18; Conservative 12; Mismatches 9; Indels 1; Gaps 1;

QY 25 RKEAKTEDEKEQRRVERVLNRRRAAQSSRRKRLEVEALE 64

DB 106 RKRRKFANE--ERRIARKLNKRTAESRRRRIRKWKILE 144

RESULT 8

T00759

Hypothetical protein At2g40950 [imported] - Arabidopsis thaliana

N:Alternate names: hypothetical protein T20B5.15

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 31-Dec-2004

A:Accession: T00759; H84835

R:Rounsley, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; Kaul

submitted to the EMBL Data Library, November 1997

A:Description: Arabidopsis thaliana chromosome II BAC T20B5 genomic sequence.

A:Reference number: Z14159

A:Accession: T00759

A:Status: translated from GB/EMBL/DDBJ

A:Molecule type: DNA

A:Residues: 1-600 <ROU>

A:Cross-references: UNIPROT:O22208; UNIPARC:UPI0000179D30; EMBL:AC002409; NID:G2623294;

A:Experimental source: cultivar Columbia

R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;

M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.

eues, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J

Nature 402, 761-768, 1999

A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.

A:Reference number: A84420; MUID:20083487; PMID:10617197

A:Accession: H84835

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-600 <STO>

A:Cross-references: UNIPARC:UPI0000179D30; GB:AE002093; NID:G2623309; PIDN:AAB86455.1; G

C:Genetics:

A:Gene: T20B5.15; At2g40950

A:Map position: 2

C:Superfamily: BZIP transcription factor; fos/jun DNA-binding domain homology

Query Match 26.9%; Score 87; DB 2; Length 600;

Best Local Similarity 32.3%; Pred. No. 1.3;

Matches 21; Conservative 16; Mismatches 22; Indels 6; Gaps 1;

QY 6 KKKKSGQVLPPEKTNLPKRKAKTED-----EKQRRVERVLNRRRAAQSSRRKRLE 59

DB 193 KRKTEIDLTDSRNSKYRRSGEDADASAVTGEDEKKRLMRNRESAQLSRQKGY 252

QY 60 VEALE 64

DB 253 VEELE 257

RESULT 9

T03373

Probable G-box binding factor 8 - rice

C:Species: Oryza sativa (rice)

C:Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 31-Dec-2004
C:Accession: T03373

R:Nakagawa, H.; Ohmiya, K.; Hattori, T.

Plant J. 9, 217-227, 1996

A:Title: A rice bZIP protein, designated OSB28, is rapidly induced by abscisic acid.

A:Reference number: Z14906; MUID:96417817; PMID:8820608

A:Accession: T03373

A:Status: preliminary; translated from GB/EMBL/DDBJ

A:Molecule type: mRNA

A:Residues: 1-360 <NAK>

A:Cross-references: UNIPROT:Q40645; UNIPARC:UPI00000A8257; EMBL:U42208; NID:gl147631; PII

A:Experimental source: cv. Nipponbare

C:Superfamily: BZIP G-box-binding protein; fos/jun DNA-binding domain homology

Query Match 26.6%; Score 86; DB 2; Length 360;

Best Local Similarity 37.7%; Pred. No. 0.98;

Matches 23; Conservative 9; Mismatches 23; Indels 6; Gaps 1;

QY 3 KPVKKKSGQVLPPEKTNLPKRKAKTEDEKEQRRVERVLNRRRAAQSSRRKRLEVEA 62

DB 199 KPDVSTASDFRVIAIPVTEVP-----TKDDKSKRRKQSNRESARRSLRKQAETEE 252

QY 63 L 63

DB 253 L 253

RESULT 10

T50972

Probable zootin [imported] - Neurospora crassa

N:Alternate names: protein B24P7.270

C:Species: Neurospora crassa

C:Date: 21-Jul-2000 #sequence_revision 21-Jul-2000 #text_change 09-Jul-2004

C:Accession: T50972

R:Schulte, U.; Aign, V.; Hoheisel, J.; Brandt, P.; Fartmann, B.; Holland, R.; Nyakatura,

submitted to the Protein Sequence Database, July 2000

A:Reference number: Z25286

A:Accession: T50972

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-445 <SCH>

A:Cross-references: UNIPROT:Q9P308; UNIPARC:UPI000006B021; EMBL:AL389890; GSPDB:GN00116;

A:Experimental source: BAC clone B24P7; strain OR74A

C:Genetics:

A:Gene: NCSP:B24P7.270

A:Map position: 6

A:Introns: 98/3

Query Match 25.1%; Score 81; DB 2; Length 445;

Best Local Similarity 35.2%; Pred. No. 3.3;

Matches 19; Conservative 11; Mismatches 16; Indels 8; Gaps 1;

QY 19 KTNLPKRKAKTEDEKEQRRV-----ERVLRNRRRAAQSSRRKRLEVEALE 64

DB 260 RKNLNTKKKAEADNARLKLDDCSAADERIKKFKQEAANAANKKKRLEKEAAE 313

RESULT 11

JC4028

activating transcription factor 2 - African clawed frog

N:Alternate names: cyclic AMP-response element-binding protein

C:Species: Xenopus laevis (African clawed frog)

C:Date: 13-Jun-1995 #sequence_revision 14-Jul-1995 #text_change 09-Jul-2004

C:Accession: JC4028

R:Villarreal, X.C.; Richter, J.D.

Gene 153, 225-229, 1995

A:Title: Analysis of ATF2 gene expression during early Xenopus laevis developme nt.

A:Reference number: JC4028; MUID:95180723; PMID:7875593

A:Accession: JC4028

A:Molecule type: mRNA

A:Residues: 1-486 <VIL>

A:Cross-references: UNIPROT:Q91576; UNIPARC:UPI00000FB2FD; GB:U16158; NID:9887779; PIDN:J

C:Comment: This protein is a sequence-specific DNA-binding protein that mediates transcri

C; Superfamily: cAMP response element-binding protein 1; fos/jun DNA-binding domain homolog
C; Keywords: leucine zipper; phosphoprotein; transcription regulation
F; 37-398/Region: leucine zipper motif
F; 329-369/Domain: fos/jun DNA-binding domain homology <FJD>
F; 332-364/Region: basic
F; 82/Binding site: phosphate (Ser) (covalent) (by protein kinase A) #status predicted
F; 102/Binding site: phosphate (Ser) (covalent) (by protein kinase C) #status predicted

Query Match 24.9%; Score 80.5; DB 2; Length 486;
Best Local Similarity 33.3%; Pred. No. 4;
Matches 19; Conservative 16; Mismatches 21; Indels 1; Gaps 1;

QY 8 RKSNGVQLPEPKTNLPKRKAETDESKBQRVERVLNRRAAQSSRRKRLVEALE 64
| | : : : || : : : || : : : || : : : || : : : ||
Db 308 RPSPAQTPTQSTSGRRRAANED-PDEKRKIIQNRAAAGRCRQKRKVWQSLE 363

RESULT 12
myosin X - bovine
C; Species: Bos primigenius taurus (cattle)
C; Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C; Accession: Tl8519
R; Corey, D.P.; Derfler, B.H.; Solc, C.K.; Duyk, G.M.; Cheney, R.E.
submitted to the EMBL Data Library, April 1996
A; Description: Cloning and expression of myosin X, a novel unconventional myosin with pl
A; Reference number: Z18942
A; Accession: Tl8519
A; Status: preliminary; translated from GB/EMBL/DDBJ
A; Molecule type: mRNA
A; Residues: 1-2052 <COR>
A; Cross-references: UNIPROT:P79114; UNIPARC:UPI000012PABC; EMBL:U55042; NID:g1755048; P
A; Experimental source: aorta
C; Keywords: nucleotide binding; P-loop
F; 66-727/Domain: myosin motor domain homology <MMO>
F; 157-164/Region: nucleotide-binding motif A (P-loop)

Query Match 24.6%; Score 79.5; DB 2; Length 2052;
Best Local Similarity 35.5%; Pred. No. 18;
Matches 22; Conservative 10; Mismatches 25; Indels 5; Gaps 1;

QY 8 RKSNGVQLPEPKTNLPKRK-----AKTEDEKQRVRVERVLNRRAAQSSRRKRLVEEA 62
| : : : | : : : || : : : || : : : || : : : || : : : ||
Db 805 RVRYVQLLAEKRAEEKKREEEKRKREEREREREALRAQQEAAKQRELEA 864

QY 63 LE 64
| :
Db 865 LQ 866

RESULT 13
S26812
transcription factor ATF-4 - mouse
N; Alternate names: activating transcription factor 4; TAXREB87 homolog; transcription fa
C; Species: Mus musculus (house mouse)
C; Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 05-Oct-2004
C; Accession: S26812; S18719; C46132; A58861; A47443
R; Pruitt, S.C.
submitted to the EMBL Data Library, August 1991
A; Reference number: S26812
A; Accession: S26812
A; Molecule type: DNA
A; Residues: 1-381 <PRU>
A; Cross-references: UNIPROT:Q61328; UNIPARC:UPI00000283A5; EMBL:X61507; NID:g50049; PIDN
R; Mielnicki, L.M.; Pruitt, S.C.
Nucleic Acids Res. 19, 6332, 1991
A; Title: Isolation and nucleotide sequence of a murine cDNA homologous to human activati
A; Reference number: S18719; MUID:92066493; PMID:1956797
A; Accession: S18719
A; Molecule type: DNA
A; Residues: 1-57,59-288,'R',290-381 <MIE>
A; Cross-references: UNIPARC:UPI0000179B78; EMBL:X61507
R; Chevray, P.M.; Nathans, D.

```

RESULT 15
S42392
G-box-binding protein - tomato
C:Species: Lycopersicon esculentum (tomato)
C>Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 31-Dec-2004
C:Accession: S42392
R:Meier, I.; Gruissem, W.
Nucleic Acids Res. 22, 470-478, 1994
A:Title: Novel conserved sequence motifs in plant G-box binding proteins and implication
A:Reference number: S42392; MUID:94173701; PMID:8127687
A:Accession: S42392
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-232 <ME1>
A:Cross-references: UNIPROT:Q43507; UNIPARC:UPI00000ACFBF; EMBL:X74941; NID:g456750; PID
C:Superfamily: BZIP G-box-binding protein; fos/jun DNA-binding domain homology
F:130-170/Domain: fos/jun DNA-binding domain homology <FJD>

Query Match      24.3%; Score 78.5; DB 2; Length 232;
Best Local Similarity 35.0%; Pred. No. 3.1;
Matches 21; Conservative 10; Mismatches 24; Indels 5; Gaps 1;

Oy      5 VKKRKSGQVLPEPTNLPPRKRAKTEDEKEQRRVERVLNRRAQSSRRERKRLVEALE 64
Db      110 IKWRNQGSGVSPAGWG-----REWIQDERELKQKQKQSNRESARRSLRKQACEELQ 164

Search completed: November 22, 2005, 15:49:29
Job time : 26 secs

```

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: November 22, 2005, 15:31:54 ; Search time 159 Seconds
(without alignments)
176.857 Million cell updates/sec

Title: US-10-663-450-5_COPY_84_147

Perfect score: 323

Sequence: 1 EKKPVKKRSGQVLEPPT.....NRRAAQSSRRKRLEVEALS 64

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_21.*

1: Geneseqp1980s.*

2: Geneseqp1990s.*

3: Geneseqp2000s.*

4: Geneseqp2001s.*

5: Geneseqp2002s.*

6: Geneseqp2003as.*

7: Geneseqp2003bs.*

8: Geneseqp2004s.*

9: Geneseqp2005s.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	323	100.0	64	5	AAE15373
2	323	100.0	450	4	AAE15373
3	323	100.0	451	5	AAE15371
4	323	100.0	451	8	ADS12801
5	323	100.0	451	8	ADS12804
6	307	95.0	409	7	ABO43144 A. thalia
7	307	95.0	409	7	ADB31925
8	307	95.0	409	8	ADO02271
9	292	90.4	342	4	AAE15377
10	292	90.4	342	5	AAE15381
11	292	90.4	342	8	ADS12818
12	292	90.4	366	5	AAE15379
13	292	90.4	386	8	ADS12815
14	288	89.2	64	5	AAE15374
15	288	89.2	349	8	ADS12805
16	288	89.2	349	8	ADS12803
17	288	89.2	349	8	ADS12863
18	288	89.2	350	4	AAE15376
19	288	89.2	350	5	AAE15372
20	175	54.2	68	8	AAE15382
21	175	54.2	68	8	ADS12859
22	175	54.2	84	5	ABP02534
23	175	54.2	200	8	ADS43437
24	175	54.2	230	2	AAE153806

25	175	54.2	230	8	ADT87049
26	175	54.2	238	2	AAW53807
27	108	33.4	192	4	AAE15371
28	108	33.4	192	4	AAE15371
29	108	33.4	192	4	AAE15371
30	108	33.4	192	4	AAE15371
31	108	33.4	192	4	AAE15371
32	106	32.8	168	5	AAU93013
33	106	32.8	168	7	ADD30174
34	106	32.8	168	8	ADI43893
35	106	32.8	211	3	AAE15371
36	103	31.9	170	8	ADM48147
37	101	31.3	185	8	ADT87049
38	100.5	31.1	185	9	ADT87049
39	99	30.7	143	3	AAE15371
40	99	30.7	163	9	ADM17580
41	94	29.1	646	7	ABM86095
42	93	28.8	120	3	AAE15371
43	93	28.8	135	3	AAE15371
44	93	28.8	149	3	AAE15371
45	93	28.8	149	3	AAE15371

ALIGNMENTS

RESULT 1

AAE15373

ID AAE15373 standard; protein; 64 AA.

XX AAE15373;

XX AAE15373;

DT 29-AUG-2003 (revised)

DT 07-MAR-2002 (first entry)

XX AAE15373;

DE Trichoderma reesei HAC1 protein DNA binding domain.

XX AAE15373;

XX AAE15373;

XX AAE15373;

XX AAE15373;

XX AAE15373;

XX AAE15373;

XX AAE15373;

XX AAE15373;

XX AAE15373;

XX AAE15373;

XX AAE15373;

XX AAE15373;

XX AAE15373;

XX AAE15373;

XX AAE15373;

XX AAE15373;

XX AAE15373;

XX AAE15373;

XX AAE15373;

XX AAE15373;

XX AAE15373;

XX AAE15373;

XX AAE15373;

XX AAE15373;

XX AAE15373;

XX AAE15373;

XX AAE15373;

XX AAE15373;

XX AAE15373;

XX AAE15373;

XX AAE15373;

XX AAE15373;

XX AAE15373;

XX AAE15373;

XX AAE15373;

XX AAE15373;

XX AAE15373;

XX AAE15373;

XX AAE15373;

XX AAE15373;

XX AAE15373;

XX AAE15373;

```
CC field)
XX
SQ Sequence 64 AA;
    Query Match      100.0%; Score 323; DB 5; Length 64;
    Best Local Similarity 100.0%; Pred. No. 1.5e-29;
    Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EKKPVKRSWGQVLPEPKTNLPPrKRKTEDEKEQRRVERVLNRRAAQSRRKRLEV 60
    |||||
Db 1 EKKPVKRSWGQVLPEPKTNLPPrKRKTEDEKEQRRVERVLNRRAAQSRRKRLEV 60
    |||||
QY 61 EALE 64
    ||||
Db 61 EALE 64

RESULT 2
AAB82975
ID AAB82975 standard; protein; 450 AA.
XX
AC AAB82975;
XX
DT 11-SEP-2003 (revised)
DT 21-DEC-2001 (first entry)
XX
DE Trichoderma reesei HAC1, involved in unfolded protein response.
XX
KW HAC1; transcription factor; unfolded protein response; protein secretion.
XX
OS Hypocrea jecorina.
XX
FH Key Location/Qualifiers
FT Domain 84..147
FT /label= DNA binding domain
XX
PN WO200172783-A2.
XX
PD 04-OCT-2001.
XX
PF 23-MAR-2001; 2001WO-US009401.
XX
PR 24-MAR-2000; 2000US-00534692.
XX
PA (GENV ) GENENCOR INT INC.
XX
PI Penttila ME, Ward M, Wang H, Valkonen MJ, Saloheimo MLA;
XX
DR WPI; 2001-626252/72.
DR N-PSDB; AAB26931.
XX
PT Increasing secretion of heterologous proteins e.g. lipase and cellulase
PT in eukaryotic cells useful in industry to increase production and
PT facilitate purification, by inducing an elevated unfolded protein
PT response.
XX
PS Claim 54; Fig 7A-B; 89pp; English.
XX
CC The present sequence is that of the HAC1 protein of Trichoderma reesei,
CC as deduced from the newly isolated HAC1 gene (see AAB26931). HAC1 protein
CC is a transcription factor involved in the unfolded protein response
CC (UPR). The invention provides methods for increasing the secretion of a
CC heterologous protein in a cell by inducing an elevated UPR. This can be
CC achieved by modulating the activity of HAC1, PTC2 or IRE1 in the cell,
CC e.g. by gene overexpression. The cell from which the protein is secreted
CC can be any cell having an UPR, such as mammalian cells, insect cells,
CC yeast and filamentous fungi. The protein of interest can be any secreted
CC protein such as a therapeutic protein or an industrial enzyme, e.g.
CC lipase, cellulase, endoglucanase-H, protease, carbohydrase, reductase,
CC oxidase, isomerase, transferase, kinase, phosphatase, alpha-amylase,
CC glucosylase, lignocellulose hemicellulase, pectinase and ligninase
CC (claimed). (Updated on 11-SEP-2003 to standardise OS field)
XX

SQ Sequence 450 AA;
    Query Match      100.0%; Score 323; DB 4; Length 450;
    Best Local Similarity 100.0%; Pred. No. 1.3e-28;
    Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EKKPVKRSWGQVLPEPKTNLPPrKRKTEDEKEQRRVERVLNRRAAQSRRKRLEV 60
    |||||
Db 84 EKKPVKRSWGQVLPEPKTNLPPrKRKTEDEKEQRRVERVLNRRAAQSRRKRLEV 143
    |||||
QY 61 EALE 64
    ||||
Db 144 EALE 147

RESULT 3
AAE15371
ID AAE15371 standard; protein; 451 AA.
XX
AC AAE15371;
XX
DT 29-AUG-2003 (revised)
DT 07-MAR-2002 (first entry)
XX
DE Trichoderma reesei HAC1 protein.
XX
KW Heterologous protein secretion; unfolded protein response; UPR; lipase;
KW cellulase; carbohydrase; industry; purification; HAC1 protein.
XX
OS Hypocrea jecorina.
XX
FH Key Location/Qualifiers
FT Binding-site 84..147
FT /label= DNA-binding_domain
XX
PN US2001034045-A1.
XX
PD 25-OCT-2001.
XX
PF 23-MAR-2001; 2001US-00816277.
XX
PR 24-MAR-2000; 2000US-00534692.
XX
PA (GENV ) GENENCOR INT INC.
XX
PI Penttila ME, Ward M, Wang H, Valkonen MJ, Saloheimo MLA;
XX
DR WPI; 2002-033728/04.
DR N-PSDB; AAD24595.
XX
PT Increasing secretion of heterologous proteins e.g. lipase and cellulase
PT in eukaryotic cells useful in industry to increase production and
PT facilitate purification, by inducing an elevated unfolded protein
PT response.
XX
PS Claim 39; Fig 7; 56pp; English.
XX
CC The present invention relates to methods for increasing the secretion of
CC heterologous protein in eukaryotic cells by inducing an elevated unfolded
CC protein response (UPR). The method involves inducing the elevated UPR by
CC increasing the presence of proteins such as HAC1, HACA, PTC2 or IRE1 in
CC cells. The method and sequences are useful for increasing the secretion
CC of heterologous proteins (e.g. lipase, cellulase, carbohydrase) in
CC eukaryotic cells useful in industry to increase protein yields and to
CC facilitate purification. The present sequence is Trichoderma reesei HAC1
CC protein. (Updated on 29-AUG-2003 to standardise OS field)
XX
SQ Sequence 451 AA;
    Query Match      100.0%; Score 323; DB 5; Length 451;
    Best Local Similarity 100.0%; Pred. No. 1.3e-28;
    Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

QY 1 EKKPVKRSWGVLPEPKTNLPKRKAKTEDEKEQRRVERVLNRRRAAQSSRRKRLEV 60
DB 84 EKKPVKRSWGVLPEPKTNLPKRKAKTEDEKEQRRVERVLNRRRAAQSSRRKRLEV 143
QY 61 EALE 64
DB 144 EALE 147

RESULT 4
ADS12801
ID ADS12801 standard; protein; 451 AA.
AC ADS12801;
XX
DT 16-DEC-2004 (first entry)
DE Trichoderma reesei hac1 chaperone and foldase.
XX
KW unfolded protein response; UPR; HAC1; PTC2;
KW unfolded protein response modulator; enzyme production; hac1; chaperone;
KW foldase; enzyme.
XX
OS Hypocrea jecorina.
XX
PN US2004186070-A1.
XX
PD 23-SEP-2004.
XX
PF 15-SEP-2003; 2003US-00663450.
XX
PR 24-MAR-2000; 2000US-00534692.
PR 23-MAR-2001; 2001US-00816277.
XX
PA (GEMV) GENENCOR INT INC.
XX
PI Penttila ME, Ward M, Wang H, Valkonen MJ, Saloheimo MLA;
XX
DR WPI; 2004-707924/69.
XX
DR N-PSDB; ADS12800.
XX
PT Increasing the secretion of a heterologous protein, such as a therapeutic
PT or an industrial enzyme, in genetically modified eukaryotic cells by
PT inducing an elevated unfolded protein response (UPR).
XX
PS Example 3; SEQ ID NO 2; 83pp; English.
XX
CC The invention describes a method of increasing the secretion of a
CC heterologous protein in a eukaryotic cell, comprising inducing an
CC elevated unfolded protein response (UPR). Also described are: an isolated
CC nucleic acid encoding a HAC1 protein, where the HAC1 protein induces
CC unfolded protein response and has less than 50% similarity to yeast HAC1
CC protein; an isolated nucleic acid encoding a HAC1 protein, where the HAC1
CC protein induces unfolded protein response and where the HAC1 protein
CC comprises a DNA binding region that has greater than 70% similarity to
CC the DNA binding region of filamentous fungi HAC1 protein; a protein
CC having unfolded protein response inducing activity and having greater
CC than 70% similarity to a fully defined amino acid sequence of 451, 349 or
CC 386, or 504 or 1232 amino acids (SEQ ID NO: 2, 4 or 16, or 12 or 14,
CC respectively) as given in the specification; a protein having an amino
CC acid sequence of SEQ ID NO: 2, 4, 8, 10, 12, 14 or 16; an isolated
CC nucleic acid encoding a PTC2 protein that modulates unfolded protein
CC response and has at least 70% similarity to a fully defined amino acid
CC sequence of 130 or 438 amino acids (SEQ ID NO: 8 or 10) as given in the
CC specification; an isolated nucleic acid encoding a IRE1 protein that
CC modulates unfolded protein response and has at least 60% similarity to a
CC fully defined amino acid sequence with SEQ ID NO: 12 or 14; a protein
CC having unfolded protein response modulating activity and having greater
CC than 70% similarity to SEQ ID NO: 8 or 10; and a cell containing a
CC heterologous nucleic acid encoding a protein having unfolded protein
CC response modulating activity and a heterologous nucleic acid encoding a
CC protein of interest to be secreted. The methods and compositions of
CC genetically manipulating cells to have an elevated unfolded protein

CC response (UPR) resulting in an increased capacity to produce secreted
CC proteins, are useful in e.g. production of therapeutic or industrial
CC enzymes. This is the amino acid sequence of Trichoderma reesei HAC1
CC chaperone and foldase.
XX
SQ Sequence 451 AA;
Query Match 100.0%; Score 323; DB 8; Length 451;
Best Local Similarity 100.0%; Pred. No. 1.3e-28;
Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 EKKPVKRSWGVLPEPKTNLPKRKAKTEDEKEQRRVERVLNRRRAAQSSRRKRLEV 60
DB 84 EKKPVKRSWGVLPEPKTNLPKRKAKTEDEKEQRRVERVLNRRRAAQSSRRKRLEV 143
QY 61 EALE 64
DB 144 EALE 147

RESULT 5
ADS12804
ID ADS12804 standard; protein; 451 AA.
XX
AC ADS12804;
XX
DT 16-DEC-2004 (first entry)
DE Trichoderma reesei hac1 DNA binding domain.
XX
DE
XX
KW unfolded protein response; UPR; HAC1; PTC2;
KW unfolded protein response modulator; enzyme production; hac1; chaperone;
KW foldase; DNA binding domain.
XX
OS Hypocrea jecorina.
XX
PN US2004186070-A1.
XX
PD 23-SEP-2004.
XX
PF 15-SEP-2003; 2003US-00663450.
XX
PR 24-MAR-2000; 2000US-00534692.
PR 23-MAR-2001; 2001US-00816277.
XX
PA (GEMV) GENENCOR INT INC.
XX
PI Penttila ME, Ward M, Wang H, Valkonen MJ, Saloheimo MLA;
XX
DR WPI; 2004-707924/69.
XX
PT Increasing the secretion of a heterologous protein, such as a therapeutic
PT or an industrial enzyme, in genetically modified eukaryotic cells by
PT inducing an elevated unfolded protein response (UPR).
XX
PS Example 3; SEQ ID NO 5; 83pp; English.
XX
CC The invention describes a method of increasing the secretion of a
CC heterologous protein in a eukaryotic cell, comprising inducing an
CC elevated unfolded protein response (UPR). Also described are: an isolated
CC nucleic acid encoding a HAC1 protein, where the HAC1 protein induces
CC unfolded protein response and has less than 50% similarity to yeast HAC1
CC protein; an isolated nucleic acid encoding a HAC1 protein, where the HAC1
CC protein induces unfolded protein response and where the HAC1 protein
CC comprises a DNA binding region that has greater than 70% similarity to
CC the DNA binding region of filamentous fungi HAC1 protein; a protein
CC having unfolded protein response inducing activity and having greater
CC than 70% similarity to a fully defined amino acid sequence of 451, 349 or
CC 386, or 504 or 1232 amino acids (SEQ ID NO: 2, 4 or 16, or 12 or 14,
CC respectively) as given in the specification; a protein having an amino
CC acid sequence of SEQ ID NO: 2, 4, 8, 10, 12, 14 or 16; an isolated
CC nucleic acid encoding a PTC2 protein that modulates unfolded protein
CC response and has at least 70% similarity to a fully defined amino acid
CC sequence of 130 or 438 amino acids (SEQ ID NO: 8 or 10) as given in the
CC specification; an isolated nucleic acid encoding a IRE1 protein that
CC modulates unfolded protein response and has at least 60% similarity to a
CC fully defined amino acid sequence with SEQ ID NO: 12 or 14; a protein
CC having unfolded protein response modulating activity and having greater
CC than 70% similarity to SEQ ID NO: 8 or 10; and a cell containing a
CC heterologous nucleic acid encoding a protein having unfolded protein
CC response modulating activity and a heterologous nucleic acid encoding a
CC protein of interest to be secreted. The methods and compositions of
CC genetically manipulating cells to have an elevated unfolded protein

CC sequence of 130 or 438 amino acids (SEQ ID NO: 8 or 10) as given in the
 CC specification; an isolated nucleic acid encoding a IRE1 protein that
 CC modulates unfolded protein response and has at least 60% similarity to a
 CC fully defined amino acid sequence with SEQ ID NO: 12 or 14; a protein
 CC having unfolded protein response modulating activity and having greater
 CC than 70% similarity to SEQ ID NO: 8 or 10; and a cell containing a
 CC heterologous nucleic acid encoding a protein having unfolded protein
 CC response modulating activity and a heterologous nucleic acid encoding a
 CC protein of interest to be secreted. The methods and compositions of
 CC genetically manipulating cells to have an elevated unfolded protein
 CC response (UPR) resulting in an increased capacity to produce secreted
 CC proteins, are useful in e.g. production of therapeutic or industrial
 CC enzymes. This is the amino acid sequence of Trichoderma reesei HAC1
 CC chaperone and foldase DNA binding domain.

XX Sequence 451 AA;

Query Match 100.0%; Score 323; DB 8; Length 451;
 Best Local Similarity 100.0%; Pred. No. 1.3e-28;
 Matches 64; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 EKKPVKKRSWGQVLPEPKTNLPKRKATDEKEQQRVERVLNRRAAQSRRKRLEV 60
 DB 84 EKKPVKKRSWGQVLPEPKTNLPKRKATDEKEQQRVERVLNRRAAQSRRKRLEV 143

QY 61 EALE 64

DB 144 EALE 147

RESULT 6

ABO43144

ID ABO43144 standard; protein; 409 AA.

XX ABO43144;

DT 23-SEP-2003 (first entry)

DE A. thaliana disease tolerance transcription factor, G1034.

XX Plant; transcription factor; disease resistance; transgenic;
 KW plant breeding; pathogens resistance; pests; resistance.

OS Arabidopsis thaliana.

XX US2003046723-A1.

PN 06-MAR-2003.

XX 22-MAR-2000; 2000US-00533029.

PR 22-MAR-2000; 2000US-00533029.

XX (HEAR/) HEARD J.

PA (BROU/) BROUN P.

PA (RIE/) RIECHMANN J L.

PA (KED/) KEDDIE J.

PA (PINE/) PINEDA O.

PA (ADAM/) ADAM L.

PA (SAMA/) SAMAHA R.

PA (ZHAN/) ZHANG J.

PA (YUG/) YU G.

PA (RATC/) RATCLIFFE O.

PA (PILG/) PILGRIM M.

PA (JIAN/) JIANG C.

PA (REUB/) REUBER L.

XX Heard J, Broun P, Riechmann JL, Keddie J, Pineda O, Adam L;

PI Samaha R, Zhang J, Yu G, Ratcliffe O, Pilgrim M, Jiang C, Reuber L;

XX WPI; 2003-521768/49.

DR N-PSDB; ACD98410.

XX

PT New transgenic plants comprising a recombinant gene that alters the
 PT plant's disease tolerance or resistance, useful in plant breeding, e.g.
 PT for generating plants with improved tolerance or resistance to diseases,
 PT pests or pathogens.

XX Claim 1; Page 100-101; 124pp; English.

XX The invention relates to a transgenic plant, comprising a recombinant
 CC polynucleotide that alters the plant's disease tolerance or resistance
 CC when compared with the same trait of another plant lacking the
 CC recombinant polynucleotide. The recombinant polynucleotide comprises a
 CC nucleotide sequence, which encodes a polypeptide comprising at least 6
 CC consecutive amino acids of any of 56 transcription factor proteins
 CC appearing as ABO43093-ABO43148. Also included are altering the disease
 CC tolerance or resistance of a plant (by: (a) transforming a plant with the
 CC recombinant polynucleotide; (b) selecting the transformed plants; and (c)
 CC identifying a transformed plant with an altered disease tolerance or
 CC resistance), altering the expression levels of at least one gene in a
 CC plant by transforming the plant with the recombinant polynucleotide,
 CC altering a plant's trait (comprising: (a) providing a database sequence;
 CC (b) comparing the database sequence with the polypeptide or
 CC polynucleotide cited above; (c) selecting a database sequence that meets
 CC the selected sequence criteria; and (d) transforming the selected
 CC database sequence in the plant) and altering a plant's trait (comprising:
 CC (a) providing a test polynucleotide; (b) hybridising the test
 CC polynucleotide at low stringency with the recombinant polynucleotide
 CC cited above; and (c) transforming the hybridising test polynucleotide in
 CC a plant to alter a trait of the plant. The transgenic plant is useful in
 CC plant breeding, particularly for generating plants with improved
 CC tolerance or resistance to diseases. The plants have commercial utility
 CC for increasing tolerance or resistance to pathogens and pests. The
 CC present sequence is an Arabidopsis thaliana transcription factor of the
 CC invention

XX Sequence 409 AA;

Query Match 95.0%; Score 307; DB 7; Length 409;

Best Local Similarity 93.8%; Pred. No. 7.9e-27;

Matches 60; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 1 EKKPVKKRSWGQVLPEPKTNLPKRKATDEKEQQRVERVLNRRAAQSRRKRLEV 60

DB 66 EKKPVKKRSWGQVLPEPKTNLPKRKATDEKEQQRVERVLNRRAAQSRRKRLEV 125

QY 61 EALE 64

DB 126 EALE 129

RESULT 7

ADB31925

ID ADB31925 standard; protein; 409 AA.

XX ADB31925;

DT 04-DEC-2003 (first entry)

DE Plant (A. thaliana) transcription factor polypeptide #79.

XX Plant; transcription factor; transgenic plant; transgenic; plant trait;
 modified trait.

OS Arabidopsis thaliana.

XX US2003101481-A1.

XX 29-MAY-2003.

XX 15-NOV-2002; 2002US-00295403.

XX 22-SEP-1998; 98US-0101349P.

PR 06-OCT-1998; 98US-0103312P.

PR 17-NOV-1998; 98US-0108734P.

PR 22-DEC-1998; 98US-0113409P.
 PR 13-SEP-1999; 99US-00394519.
 XX
 PA (ZHAN/) ZHANG J.
 PA (FROM/) FROMM M.
 PA (HEAR/) HEARD J.
 PA (RIEC/) RIECHMANN J L.
 PA (ADAM/) ADAM L.
 PA (BROU/) BROUN P.
 PA (PINE/) PINEDA O.
 PA (REUB/) REUBER L.
 PA (KEDD/) KEDDIE J.
 PA (YUGG/) YU G.
 PA (JIANG/) JIANG C.
 XX
 PI Zhang J, Fromm M, Heard J, Riechmann JL, Adam L, Broun P;
 PI Pineda O, Reuber L, Keddle J, Yu G, Jiang C;
 XX
 DR WPI; 2003-597572/56.
 DR N-PSDB; ADB31924.
 XX
 PT New isolated polypeptides and polynucleotide sequences, useful for
 PT screening a molecule to identify a molecule that modifies plant trait,
 PT and for producing plants with modified traits.
 XX
 PS Claim 9; SEQ ID NO 158; 17pp; English.
 XX
 CC The present invention relates to the isolation of plant (Arabidopsis
 CC thaliana) transcription factor polypeptide and polynucleotide sequences.
 CC Also disclosed are: an expression vector comprising the isolated
 CC polynucleotide, a host cell comprising the expression vector, a
 CC transgenic plant comprising the isolated polynucleotide, a transgenic
 CC plant ectopically expressing the isolated polynucleotide or polypeptide,
 CC a method for screening a molecule to identify a molecule that modifies a
 CC plant trait by placing the molecule in contact with the plant, and
 CC monitoring the effect of the molecule on the expressing or activity of
 CC the polypeptide or polynucleotide, and producing a transgenic plant
 CC having a modified trait by ectopically expressing the isolated
 CC polypeptide and selecting a plant with the modified trait. The
 CC polypeptides, polynucleotides and methods are useful for screening a
 CC molecule to identify a molecule that modifies plant trait, and for
 CC producing plants with modified traits. The present sequence represents a
 CC plant transcription factor polypeptide of the invention. Note: The
 CC sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from the
 CC USPTO web site at seqdata.uspto.gov.
 XX
 SQ Sequence 409 AA;
 Query Match 95.0%; Score 307; DB 7; Length 409;
 Best Local Similarity 93.8%; Pred. No. 7.9e-27;
 Matches 60; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
 OY 1 EKKPVKRSWGQVLPEPTNLPPRKRAKTEKQRRVERVLRNRAAQSRRKRLEV 60
 DB 66 EKKPIKRSWGQQLPEPTNLPPRKRAKTEKQRRVERVLRNRAAQSRRKRQEV 125
 OY 61 EALE 64
 DB 126 EALE 129
 RESULT 8
 ID ADO02271 standard; protein; 409 AA.
 XX
 AC ADO02271;
 XX
 DT 01-JUL-2004 (first entry)
 XX
 DE Thalecress transcription factor protein #342.
 XX
 KW Thalecress; transcription factor; plant; transgenic; abiotic stress;

KW cold tolerance; heat tolerance; drought; osmotic stress;
 KW phosphate limitation; potassium limitation; nitrogen limitation;
 KW hormone sensitivity; disease resistance; sugar sensing; seed germination;
 KW flowering; inflorescence architectural change;
 KW meristem cell differentiation; phyllotaxy; apical dominance;
 KW trichome development; seed development; premature senescence;
 KW delayed senescence; lethality; necrosis; plant size; leaf morphology;
 KW seed morphology; secondary metabolism; light response; shade avoidance.
 XX Arabidopsis thaliana.
 OS
 XX US2004045049-A1.
 PN
 XX
 XX
 PD 04-MAR-2004.
 XX
 PF 10-APR-2003; 2003US-00412699.
 XX
 PR 13-SEP-1999; 99US-00394519.
 PR 21-JAN-2000; 2000US-00489376.
 PR 17-FEB-2000; 2000US-00506720.
 PR 22-MAR-2000; 2000US-00532591.
 PR 22-MAR-2000; 2000US-00533029.
 PR 22-MAR-2000; 2000US-00533030.
 PR 22-MAR-2000; 2000US-00533392.
 PR 22-MAR-2000; 2000US-00533648.
 PR 06-APR-2000; 2000WO-US009448.
 PR 16-NOV-2000; 2000US-00713994.
 PR 27-MAR-2001; 2001US-00819142.
 PR 17-APR-2001; 2001US-00837444.
 PR 30-JAN-2002; 2002US-00958131.
 PR 14-JUN-2002; 2002US-00171468.
 PR 09-AUG-2002; 2002US-00225066.
 PR 09-AUG-2002; 2002US-00225067.
 PR 09-AUG-2002; 2002US-00225068.
 PR 17-DEC-2002; 2002US-0434166P.
 PR 25-FEB-2003; 2003US-00374780.
 XX (ZHAN/) ZHANG J.
 PA (FROM/) FROMM M.
 PA (HEAR/) HEARD J E.
 PA (RIEC/) RIECHMANN J L.
 PA (ADAM/) ADAM L J.
 PA (BROU/) BROUN P E.
 PA (PINE/) PINEDA O.
 PA (REUB/) REUBER T L.
 PA (KEDD/) KEDDIE J S.
 PA (YUGG/) YU G.
 PA (JIAN/) JIANG C.
 PA (SAMA/) SAMAHA R S.
 PA (PILG/) PILGRIM M L.
 PA (CREE/) CREELMAN R A.
 PA (DUBE/) DUBELL A N.
 PA (RATC/) RATCLIFFE O.
 PA (KUMI/) KUMIMOTO R.
 PA (SHER/) SHERMAN B K.
 XX
 PI Zhang J, Fromm ME, Heard JE, Riechmann JL, Adam LJ, Broun PE;
 PI Pineda O, Reuber TL, Keddle JS, Yu G, Jiang C, Samaha RS;
 PI Pilgrim ML, Creelman RA, Dubell AN, Ratcliffe O, Kumimoto R;
 PI Sherman BK;
 XX
 XX WPI; 2004-225755/21.
 DR N-PSDB; ADO02270.
 DR
 XX
 PT New transgenic plant, useful in developing phenotypes with altered or
 PT improved characteristics or traits.
 XX
 PS Claim 1; SEQ ID NO 684; 213pp; English.
 XX
 CC The invention relates to a transgenic plant comprises a recombinant
 CC polynucleotide having a polynucleotide sequence or its complementary
 CC sequence comprising a sequence encoding a polypeptide, that initiates
 CC transcription (i.e. a transcription factor) from Arabidopsis, Soybean,

cellulase; carbohydrase; industry; purification; hacA protein.
 Aspergillus niger.
 US2001034045-A1.
 25-OCT-2001.
 23-MAR-2001; 2001US-00816277.
 24-MAR-2000; 2000US-00534692.
 (GEMV) GENENCOR INT INC.
 Penttila ME, Ward M, Wang H, Valkonen MJ, Saloheimo MLA;
 WPI; 2002-033728/04.
 Increasing secretion of heterologous proteins e.g. lipase and cellulase
 in eukaryotic cells useful in industry to increase production and
 facilitate purification, by inducing an elevated unfolded protein
 response.
 Claim 39; Fig 28; 56pp; English.
 The present invention relates to methods for increasing the secretion of
 heterologous protein in eukaryotic cells by inducing an elevated unfolded
 protein response (UPR). The method involves inducing the elevated UPR by
 increasing the presence of proteins such as HAC1, HACA, PTC2 or IRE1 in
 cells. The method and sequences are useful for increasing the secretion
 of heterologous proteins (e.g. lipase, cellulase, carbohydrase) in
 eukaryotic cells useful in industry to increase protein yields and to
 facilitate purification. The present sequence is Aspergillus niger var.
 awamori hacA protein
 Sequence 342 AA;
 Query Match 90.4%; Score 292; DB 5; Length 342;
 Best Local Similarity 89.1%; Pred. No. 3.5e-25;
 Matches 57; Conservative 3; Mismatches 4; Indels 0; Gaps 0;
 QY 1 EKKPVKKRSGQVLPEPKTNLPKRKRAKTEDEKEQRRVERVLRNRRAAQSRRKRLEV 60
 DB 46 EKKPVKKRSGQQLPVPKTNLPKRKRAKTEDEKEQRRVERVLRNRRAAQSRRKRLEM 105
 QY 61 EALE 64
 DB 106 EKLE 109
 RESULT 11
 ADS12818
 ID ADS12818 standard; protein; 342 AA.
 AC ADS12818;
 XX 16-DEC-2004 (first entry)
 XX Aspergillus nidulans hacA polypeptide seqid 19.
 DE unfolded protein response; UPR; HAC1; PTC2;
 KW unfolded protein response modulator; enzyme production; hacA; chaperone;
 KW foldase.
 XX Emericella nidulans.
 OS US2004186070-A1.
 PN 23-SEP-2004.
 XX 15-SEP-2003; 2003US-00663450.
 PF 24-MAR-2000; 2000US-00534692.
 PR

23-MAR-2001; 2001US-00816277.
 (GEMV) GENENCOR INT INC.
 Penttila ME, Ward M, Wang H, Valkonen MJ, Saloheimo MLA;
 WPI; 2004-707924/69.
 Increasing the secretion of a heterologous protein, such as a therapeutic
 or an industrial enzyme, in genetically modified eukaryotic cells by
 inducing an elevated unfolded protein response (UPR).
 Example 12; SEQ ID NO 19; 83pp; English.
 The invention describes a method of increasing the secretion of a
 heterologous protein in a eukaryotic cell, comprising inducing an
 elevated unfolded protein response (UPR). Also described are: an isolated
 nucleic acid encoding a HAC1 protein, where the HAC1 protein induces
 unfolded protein response and has less than 50% similarity to yeast HAC1
 protein; an isolated nucleic acid encoding a HAC1 protein, where the HAC1
 protein induces unfolded protein response and where the HAC1 protein
 comprises a DNA binding region that has greater than 70% similarity to
 the DNA binding region of filamentous fungi HAC1 protein; a protein
 having unfolded protein response inducing activity and having greater
 than 70% similarity to a fully defined amino acid sequence of 451, 349 or
 386, or 504 or 1232 amino acids (SEQ ID NO: 2, 4 or 16, or 12 or 14,
 respectively) as given in the specification; a protein having an amino
 acid sequence of SEQ ID NO: 2, 4, 8, 10, 12, 14 or 16; an isolated
 nucleic acid encoding a PTC2 protein that modulates unfolded protein
 response and has at least 70% similarity to a fully defined amino acid
 sequence of 130 or 438 amino acids (SEQ ID NO: 8 or 10) as given in the
 specification; an isolated nucleic acid encoding a IRE1 protein that
 modulates unfolded protein response and has at least 60% similarity to a
 fully defined amino acid sequence with SEQ ID NO: 12 or 14; a protein
 having unfolded protein response modulating activity and having greater
 than 70% similarity to SEQ ID NO: 8 or 10; and a cell containing a
 heterologous nucleic acid encoding a protein having unfolded protein
 response modulating activity and a heterologous nucleic acid encoding a
 protein of interest to be secreted. The methods and compositions of
 genetically manipulating cells to have an elevated unfolded protein
 response (UPR) resulting in an increased capacity to produce secreted
 proteins, are useful in e.g. production of therapeutic or industrial
 enzymes. This is the amino acid sequence of an Aspergillus nidulans hacA
 chaperone and foldase polypeptide.
 Sequence 342 AA;
 Query Match 90.4%; Score 292; DB 8; Length 342;
 Best Local Similarity 89.1%; Pred. No. 3.5e-25;
 Matches 57; Conservative 3; Mismatches 4; Indels 0; Gaps 0;
 QY 1 EKKPVKKRSGQVLPEPKTNLPKRKRAKTEDEKEQRRVERVLRNRRAAQSRRKRLEV 60
 DB 46 EKKPVKKRSGQQLPVPKTNLPKRKRAKTEDEKEQRRVERVLRNRRAAQSRRKRLEM 105
 QY 61 EALE 64
 DB 106 EKLE 109
 RESULT 12
 AAE15379
 ID AAE15379 standard; protein; 386 AA.
 AC AAE15379;
 XX 07-MAR-2002 (first entry)
 XX Aspergillus niger var. awamori hacA protein #1.
 DE Heterologous protein secretion; unfolded protein response; UPR; lipase;
 KW cellulase; carbohydrase; industry; purification; hacA protein.
 XX

```
OS Aspergillus niger.
PN US2001034045-A1.
XX
PD 25-OCT-2001.
XX
PF 23-MAR-2001; 2001US-00816277.
XX
PR 24-MAR-2000; 2000US-00534692.
XX
PA (GEMV ) GENENCOR INT INC.
XX
PI Penttila ME, Ward M, Wang H, Valkonen MJ, Saloheimo MLA;
XX
DR WPI; 2002-033728/04.
DR N-PSDB; AAD24601.
XX
PT Increasing secretion of heterologous proteins e.g. lipase and cellulase
PT in eukaryotic cells useful in industry to increase production and
PT facilitate purification, by inducing an elevated unfolded protein
PT response.
XX
PS Claim 39; Fig 28; 56pp; English.
XX
CC The present invention relates to methods for increasing the secretion of
CC heterologous protein in eukaryotic cells by inducing an elevated unfolded
CC protein response (UPR). The method involves inducing the elevated UPR by
CC increasing the presence of proteins such as HAC1, HAC2, PTC2 or lrel1 in
CC cells. The method and sequences are useful for increasing the secretion
CC of heterologous proteins (e.g. lipase, cellulase, chaperone) in
CC eukaryotic cells useful in industry to increase protein yields and to
CC facilitate purification. The present sequence is Aspergillus niger var.
CC awamori hacA protein
XX
SQ Sequence 386 AA;
Query Match 90.4%; Score 292; DB 5; Length 386;
Best Local Similarity 89.1%; Pred. No. 4e-25;
Matches 57; Conservative 3; Mismatches 4; Indels 0; Gaps 0;
QY 1 EKKPVKGRKSWGQVLPEPKTNLPKRKRAKTEDEKQRRVERVLNRRAAQSRRKRLEV 60
DB 90 EKKPVKGRKSWGQELVPKTNLPKRKRAKTEDEKQRRVERVLNRRAAQSRRKRLEM 149
QY 61 EALE 64
DB 150 EKLE 153
RESULT 13
ADSI2815 standard; protein; 386 AA.
XX
AC ADSI2815;
XX
DT 16-DEC-2004 (first entry)
XX
DE Aspergillus nidulans hacA chaperone and foldase #3.
XX
KW unfolded protein response; UPR; HAC1; PTC2;
KW unfolded protein response modulator; enzyme production; hacA; chaperone;
KW foldase; enzyme.
XX
OS Emericella nidulans.
XX
PN US2004186070-A1.
XX
PD 23-SEP-2004.
XX
PF 15-SEP-2003; 2003US-00663450.
XX
PR 24-MAR-2000; 2000US-00534692.
PR 23-MAR-2001; 2001US-00816277.
```

```
XX (GEMV ) GENENCOR INT INC.
XX
XX Penttila ME, Ward M, Wang H, Valkonen MJ, Saloheimo MLA;
XX
XX WPI; 2004-707924/69.
XX N-PSDB; ADSI2814.
XX
XX Increasing the secretion of a heterologous protein, such as a therapeutic
XX or an industrial enzyme, in genetically modified eukaryotic cells by
XX inducing an elevated unfolded protein response (UPR).
XX
XX Example 12; SEQ ID NO 16; 83pp; English.
XX
XX The invention describes a method of increasing the secretion of a
XX heterologous protein in a eukaryotic cell, comprising inducing an
XX elevated unfolded protein response (UPR). Also described are: an isolated
XX nucleic acid encoding a HAC1 protein, where the HAC1 protein induces
XX unfolded protein response and has less than 50% similarity to yeast HAC1
XX protein; an isolated nucleic acid encoding a HAC1 protein, where the HAC1
XX protein induces unfolded protein response and where the HAC1 protein
XX comprises a DNA binding region that has greater than 70% similarity to
XX the DNA binding region of filamentous fungi HAC1 protein; a protein
XX having unfolded protein response inducing activity and having greater
XX than 70% similarity to a fully defined amino acid sequence of 451, 349 or
XX 386, or 504 or 1232 amino acids (SEQ ID NO: 2, 4 or 16, or 12 or 14,
XX respectively) as given in the specification; a protein having an amino
XX acid sequence of SEQ ID NO: 2, 4, 8, 10, 12, 14 or 16; an isolated
XX nucleic acid encoding a PTC2 protein that modulates unfolded protein
XX response and has at least 70% similarity to a fully defined amino acid
XX sequence of 130 or 438 amino acids (SEQ ID NO: 8 or 10) as given in the
XX specification; an isolated nucleic acid encoding a lrel1 protein that
XX modulates unfolded protein response and has at least 60% similarity to a
XX fully defined amino acid sequence with SEQ ID NO: 12 or 14; a protein
XX having unfolded protein response modulating activity and having greater
XX than 70% similarity to SEQ ID NO: 8 or 10; and a cell containing a
XX heterologous nucleic acid encoding a protein having unfolded protein
XX response modulating activity and a heterologous nucleic acid encoding a
XX protein of interest to be secreted. The methods and compositions of
XX genetically manipulating cells to have an elevated unfolded protein
XX response (UPR) resulting in an increased capacity to produce secreted
XX proteins, are useful in e.g. production of therapeutic or industrial
XX enzymes. This is the amino acid sequence of Aspergillus nidulans HAC1
XX chaperone and foldase.
XX
XX Sequence 386 AA;
Query Match 90.4%; Score 292; DB 8; Length 386;
Best Local Similarity 89.1%; Pred. No. 4e-25;
Matches 57; Conservative 3; Mismatches 4; Indels 0; Gaps 0;
QY 1 EKKPVKGRKSWGQVLPEPKTNLPKRKRAKTEDEKQRRVERVLNRRAAQSRRKRLEV 60
DB 90 EKKPVKGRKSWGQELVPKTNLPKRKRAKTEDEKQRRVERVLNRRAAQSRRKRLEM 149
QY 61 EALE 64
DB 150 EKLE 153
RESULT 14
AAEI5374
ID AAEI5374 standard; protein; 64 AA.
XX
XX AAEI5374;
XX
XX 29-AUG-2003 (revised)
XX 07-MAR-2002 (first entry)
XX
XX Aspergillus nidulans hacA protein DNA binding domain.
XX
XX Heterologous protein secretion; unfolded protein response; UPR; lipase;
XX cellulase; carbohydrase; industry; purification; DNA binding domain;
KW
```

hacA protein.
 Emericella nidulans.
 US2001034045-A1.
 25-OCT-2001.
 23-MAR-2001; 2001US-00816277.
 24-MAR-2000; 2000US-00534692.
 (GEMV) GENENCOR INT INC.
 Penttila ME, Ward M, Wang H, Valkonen MJ, Saloheimo MLA;
 WPI; 2002-033728/04.
 Increasing secretion of heterologous proteins e.g. lipase and cellulase
 in eukaryotic cells useful in industry to increase production and
 facilitate purification, by inducing an elevated unfolded protein
 response.
 Example 3; Fig 10; 56pp; English.
 The present invention relates to methods for increasing the secretion of
 heterologous protein in eukaryotic cells by inducing an elevated unfolded
 protein response (UPR). The method involves inducing the elevated UPR by
 increasing the presence of proteins such as HAC1, HACA, PTC2 or IRE1 in
 cells. The method and sequences are useful for increasing the secretion
 of heterologous proteins (e.g. lipase, cellulase, carbohydrase) in
 eukaryotic cells useful in industry to increase protein yields and to
 facilitate purification. The present sequence is Aspergillus nidulans
 hacA protein DNA binding domain. (Updated on 29-AUG-2003 to standardise
 OS field)
 Sequence 64 AA;
 Query Match 89.2%; Score 288; DB 5; Length 64;
 Best Local Similarity 87.5%; Pred. No. 1.6e-25;
 Matches 56; Conservative 3; Mismatches 5; Indels 0; Gaps 0;
 OY 1 EKKPVKKRKSQGVLPKPTNLPPKRAKTEDEKEQRRVRLNRRRAAQSRRKRLEV 60
 DB 1 EKKPAKRRKSQGVLPKPTNLPPKRAKTEDEKEQRRRIERVLNRRRAAQTSSRRKRLEM 60
 OY 61 EALE 64
 DB 61 EKLE 64
 RESULT 15
 ADS12805
 ID ADS12805 standard; protein; 349 AA.
 XX
 AC ADS12805;
 XX
 DT 16-DEC-2004 (first entry)
 XX
 DE Aspergillus nidulans hacA DNA binding domain.
 XX
 KW unfolded protein response; UPR; HAC1; PTC2;
 KW unfolded protein response modulator; enzyme production; hacA; chaperone;
 KW foldase; DNA binding domain.
 XX
 OS Emericella nidulans.
 XX
 PN US2004186070-A1.
 XX
 PD 23-SEP-2004.
 XX
 PF 15-SEP-2003; 2003US-00663450.
 XX

PR 24-MAR-2000; 2000US-00534692.
 PR 23-MAR-2001; 2001US-00816277.
 XX
 PA (GEMV) GENENCOR INT INC.
 XX
 PI Penttila ME, Ward M, Wang H, Valkonen MJ, Saloheimo MLA;
 XX
 XX WPI; 2004-707924/69.
 DR
 XX
 PT Increasing the secretion of a heterologous protein, such as a therapeutic
 PT or an industrial enzyme, in genetically modified eukaryotic cells by
 PT inducing an elevated unfolded protein response (UPR).
 XX
 PS Example 3; SEQ ID NO 6; 83pp; English.
 XX
 CC The invention describes a method of increasing the secretion of a
 CC heterologous protein in a eukaryotic cell, comprising inducing an
 CC elevated unfolded protein response (UPR). Also described are: an isolated
 CC nucleic acid encoding a HAC1 protein, where the HAC1 protein induces
 CC unfolded protein response and has less than 50% similarity to yeast HAC1
 CC protein; an isolated nucleic acid encoding a HAC1 protein, where the HAC1
 CC protein induces unfolded protein response and where the HAC1 protein
 CC comprises a DNA binding region that has greater than 70% similarity to
 CC the DNA binding region of filamentous fungi HAC1 protein; a protein
 CC having unfolded protein response inducing activity and having greater
 CC than 70% similarity to a fully defined amino acid sequence of 451, 349 or
 CC 386, or 504 or 1232 amino acids (SEQ ID NO: 2, 4 or 16, or 12 or 14,
 CC respectively) as given in the specification; a protein having an amino
 CC acid sequence of SEQ ID NO: 2, 4, 8, 10, 12, 14 or 16; an isolated
 CC nucleic acid encoding a PTC2 protein that modulates unfolded protein
 CC response and has at least 70% similarity to a fully defined amino acid
 CC sequence of 130 or 438 amino acids (SEQ ID NO: 8 or 10) as given in the
 CC specification; an isolated nucleic acid encoding a IRE1 protein that
 CC modulates unfolded protein response and has at least 60% similarity to a
 CC fully defined amino acid sequence with SEQ ID NO: 12 or 14; a protein
 CC having unfolded protein response modulating activity and having greater
 CC than 70% similarity to SEQ ID NO: 8 or 10; and a cell containing a
 CC heterologous nucleic acid encoding a protein having unfolded protein
 CC response modulating activity and a heterologous nucleic acid encoding a
 CC protein of interest to be secreted. The methods and compositions of
 CC genetically manipulating cells to have an elevated unfolded protein
 CC response (UPR) resulting in an increased capacity to produce secreted
 CC proteins, are useful in e.g. production of therapeutic or industrial
 CC enzymes. This is the amino acid sequence of Aspergillus nidulans HAC1
 CC chaperone and foldase DNA binding domain.
 XX
 SQ Sequence 349 AA;
 Query Match 89.2%; Score 288; DB 8; Length 349;
 Best Local Similarity 87.5%; Pred. No. 1e-24;
 Matches 56; Conservative 3; Mismatches 5; Indels 0; Gaps 0;
 OY 1 EKKPVKKRKSQGVLPKPTNLPPKRAKTEDEKEQRRVRLNRRRAAQSRRKRLEV 60
 DB 52 EKKPAKRRKSQGVLPKPTNLPPKRAKTEDEKEQRRRIERVLNRRRAAQTSSRRKRLEM 111
 OY 61 EALE 64
 DB 112 EKLE 115
 Search completed: November 22, 2005, 15:43:30
 Job time : 161.secs

THIS PAGE BLANK (USPTO)